

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau**BL**

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|  |  |   |
|--|--|---|
| <b>(51) International Patent Classification <sup>5</sup> :</b><br><b>C12N 15/12, 15/54, 9/12</b><br><b>// C12Q 1/68, C12N 15/11</b>  | <b>A1</b>  | <b>(11) International Publication Number:</b> <b>WO 93/15201</b><br><b>(43) International Publication Date:</b> <b>5 August 1993 (05.08.93)</b> |
| <b>(21) International Application Number:</b> PCT/US93/00586<br><b>(22) International Filing Date:</b> 22 January 1993 (22.01.93)<br><b>(30) Priority data:</b><br>826,935 22 January 1992 (22.01.92) US<br><b>(60) Parent Application or Grant</b><br><b>(63) Related by Continuation</b><br>US 826,935 (CIP)<br>Filed on 22 January 1992 (22.01.92)<br><b>(71) Applicant (for all designated States except US):</b> NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US). | <b>(72) Inventors; and</b><br><b>(75) Inventors/Applicants (for US only) :</b> AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).<br><b>(74) Agents:</b> GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).<br><b>(81) Designated States:</b> AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br><br><b>Published</b><br><i>With international search report.</i><br><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |   |
| <b>(54) Title:</b> NOVEL PROTEIN TYROSINE KINASES<br><br><b>(57) Abstract</b><br><br>The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.  |  |   |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |                                       |    |                          |
|----|--------------------------|----|---------------------------------------|----|--------------------------|
| AT | Austria                  | FR | France                                | MR | Mauritania               |
| AU | Australia                | GA | Gabon                                 | MW | Malawi                   |
| BB | Barbados                 | GB | United Kingdom                        | NL | Netherlands              |
| BE | Belgium                  | GN | Guinea                                | NO | Norway                   |
| BF | Burkina Faso             | GR | Greece                                | NZ | New Zealand              |
| BG | Bulgaria                 | HU | Hungary                               | PL | Poland                   |
| BJ | Benin                    | IE | Ireland                               | PT | Portugal                 |
| BR | Brazil                   | IT | Italy                                 | RO | Romania                  |
| CA | Canada                   | JP | Japan                                 | RU | Russian Federation       |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                     | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                            | SK | Slovak Republic          |
| CI | Côte d'Ivoire            | LJ | Liechtenstein                         | SN | Senegal                  |
| CM | Cameroon                 | LK | Sri Lanka                             | SU | Soviet Union             |
| CS | Czechoslovakia           | LJ | Luxembourg                            | TD | Chad                     |
| CZ | Czech Republic           | MC | Monaco                                | TG | Togo                     |
| DE | Germany                  | MG | Madagascar                            | UA | Ukraine                  |
| DK | Denmark                  | ML | Mali                                  | US | United States of America |
| ES | Spain                    | MN | Mongolia                              | VN | Viet Nam                 |
| FI | Finland                  |    |                                       |    |                          |

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and  
5 differentiation is regulated in part by phosphorylation of  
various cellular proteins. Protein tyrosine kinases are  
enzymes that catalyze this process. Moreover, many act as  
growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein  
tyrosine kinase genes present in human megakaryocytic and  
lymphocytic cells, the proteins encoded by these genes,  
antibodies specific for the encoded proteins, RNA nucleic  
acid sequences which hybridize to the genes and methods of  
15 use therefor.

The genes isolated as described herein are referred  
to, collectively, as protein tyrosine kinase (pTK) genes.  
The nucleic acid sequences of these genes, isolated as  
discussed herein, show significant homology with  
20 previously identified protein tyrosine kinases containing  
extracellular domains which function as growth factor  
receptors. The pTK genes have been shown to be present in  
both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show  
25 significant sequence homology with members of the c-kit  
subgroup of growth factor receptors with protein tyrosine  
kinase activity. The c-kit subgroup of receptor tyrosine  
kinases catalyze the phosphorylation of exogenous  
substrates, as well as tyrosine residues within their own  
30 polypeptide chains. (Ullrich, A. and Schlessinger, J.,  
Cell, 61:203 (1990)). Members of the c-kit subgroup  
include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells, has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins  
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKS)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present  
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This  
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK  
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the  
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other  
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target  
5 proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or  
10 bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect  
15 cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in  
20 extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

#### Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1  
25 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.



-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

#### Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and  
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to  
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of  
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA  
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide  
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases  
(Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and  
the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30  
5 respectively, also exhibit sequence homology with known  
protein tyrosine kinases.

Thus, as described above, DNA which hybridize with  
DNA encoding amino acid sequences present in the catalytic  
domain of a protein tyrosine kinase of the c-kit subgroup  
10 of protein kinases have been isolated and sequenced.  
These isolated DNA sequences, collectively referred to as  
pTKs genes, (and their deduced amino acid sequences) have  
been shown to exhibit significant sequence homology with  
known members of receptor tyrosine kinase families.

15 Once isolated, these DNA fragments can be amplified  
using known standard techniques such as PCR. These  
amplified fragments can then be cloned into appropriate  
cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning  
20 vectors, labeled with a radiolabeled nucleotide such as  
<sup>32</sup>P and used to screen appropriate cDNA libraries to  
obtain the full-length cDNA clone.

The pTk genes as described above have been isolated  
from the source in which they occur naturally, i.e.  
25 megakaryocyte and lymphocytic cells. The present invention  
is intended to include pTk genes produced using genetic  
engineering techniques, such as recombinant technology, as  
well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTk genes  
30 include amino acid sequences which encode peptides  
exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for  
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,  
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the  
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a  
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a  
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be  
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector  
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For  
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design  
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or  
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of  
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be  
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques  
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced  
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase  
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or  
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase  
30 activity, or activate other growth factors.

-13-

PTK genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase  
5 activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of  
10 treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the  
15 following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the  
pTK Genes

To facilitate the isolation and identification of  
20 these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2). These sequences were used as polymerase  
25 chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the  
30 catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK  
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able  
20 to ascertain using no more than routine experimentation,  
many equivalents to the specific embodiments of the  
invention described herein. Such equivalents are intended  
to be encompassed by the following claims.



-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA  
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences  
10 consisting of:
  - a) SAL-S1 (SEQ ID NOS:5 and 7);
  - b) SAL-D4 (SEQ ID NO:7);
  - c) LpTK 2 (SEQ ID NOS:9 and 19);
  - d) LpTK 3 (SEQ ID NO:11);
  - 15 e) LpTK 4 (SEQ ID NOS:13 and 21);
  - f) LpTK 13 (SEQ ID NO:15);
  - g) LpTK 25 (SEQ ID NO: 22 and
  - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid  
20 sequence selected from the group consisting of:
  - a) SAL-S1 (SEQ ID NOS:6 and 18);
  - b) SAL-D4 (SEQ ID NO:8);
  - c) LpTK 2 (SEQ ID NOS:10 and 20);
  - d) LpTK 3 (SEQ ID NO:12);
  - 25 e) LpTK 4 (SEQ ID NO:14);
  - f) LpTK 13 (SEQ ID NO:16);
  - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
  - i) bpTK 2 (SEQ ID NO:26);
  - j) bpTK 3 (SEQ ID NO:27);
  - k) bpTK 4 (SEQ ID NO:28);
  - 5 l) bpTK 5 (SEQ ID NO:29); and
  - m) bpTK 7 (SEQ ID NO:30).
4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain
- 10 of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:6);
  - 15 b) SAL-D4 (SEQ ID NO:8);
  - c) LpTK 2 (SEQ ID NO:10);
  - d) LpTK 3 (SEQ ID NO:12);
  - e) LpTK 4 (SEQ ID NO:14); and
  - f) LpTK 13 (SEQ ID NO:16).
  - 20 g) HpTK 5 (SEQ ID NO:24);
  - h) bpTK 1 (SEQ ID NO:25);
  - i) bpTK 2 (SEQ ID NO:26);
  - j) bpTK 3 (SEQ ID NO:27);
  - k) bpTK 4 (SEQ ID NO:28);
  - 25 l) bpTK 5 (SEQ ID NO:29); and
  - m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
  - a) SAL-S1 (SEQ ID NO:6);
  - b) SAL-D4 (SEQ ID NO:8);
  - 10 c) LpTK 2 (SEQ ID NO:10);
  - d) LpTK 3 (SEQ ID NO:12);
  - e) LpTk 4 (SEQ ID NO:14); and
  - f) LpTK 13 (SEQ ID NO:16).
  - g) HpTK 5 (SEQ ID NO:24);
  - 15 h) bpTK 1 (SEQ ID NO:25);
  - i) bpTK 2 (SEQ ID NO:26);
  - j) bpTK 3 (SEQ ID NO:27);
  - k) bpTK 4 (SEQ ID NO:28);
  - l) bpTK 5 (SEQ ID NO:29); and
  - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide  
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-18-

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim 10 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim 10 encoded by the amino acid sequence (SEQ ID NO:8).
16. A protein of human megakaryocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
- 15 17. A protein of Claim 14 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim 14 encoded by the amino acid sequence (SEQ ID NO:12).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:5);
  - b) SAL-D4 (SEQ ID NO:7);
  - c) LpTK 2 (SEQ ID NO:9);
  - d) LpTK 3 (SEQ ID NO:11);
  - e) LpTk 4 (SEQ ID NO:13); and
  - f) LpTK 13 (SEQ ID NO:15).
  - g) LpTK 25 (SEQ ID NO: 22 and
  - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

1/35

sals1 (160 bases)  
FLKI-LIKE

|   |   |     |              |
|---|---|-----|--------------|
| 1   | PTKI/3 PRIMERS                            | 21  | 41           |
| 5' <u>ggatcctgtgcatcagtgacttagggctagggaacattctgtctcggaagcga</u> <u>cgtggt</u> |   |     |              |
|   | D P V H Q a L R A R N I L L S E S D V V   |     |              |
| 61  |   | 81  | 101          |
| gaagatctgtgactttggccttggcccggaacatctacaagagaccagctacgtccgcaa                  |   |     |              |
|   | K I C D F G L A R D I Y K D P S (Y) V R K |     |              |
| 121   |   | 141 | PTKKW PRIMER |
| <u>gcatgcccggtgccctgaagtggatggcgccagaattc</u> 3'                              |   |     |              |
|   | H A R L P L K W H A P E F                 |     |              |

FIGURE 1

sald4 (147 bases)-  
FGFR-LIKE

|     |   |     |    |
|-----|---|-----|----|
| 1   | PTKI/3 PRIMERS  | 21  | 41 |
| 6'  | <u>ggatccattcacagagacctagcagcacgcaacatcctgggtcttcagaggacctggtaacc</u> |     |    |
|     | G S I H R D L A A R N I L V S E D L V T                               |     |    |
| 61  | 81  | 101 |    |
|     | <u>aaggtcagcgactttggcctggccaagccgagcggaagggttagactcaagccggctg</u>     |     |    |
|     | K V S D F G L A K A E R K G L D S S R L                               |     |    |
| 121 | PTKKW PRIMER  | 141 |    |
|     | <u>ccgtcaaatggatggctcccgaattc</u> 3'                                  |     |    |
|     | P V K W H A P E F   |     |    |

FIGURE 2

3/35

**LpTK2**

GTTGGAATTCCTTCGGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA  
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG  
ATATTATGTTACCAACGAGGACATTCCT

**FIGURE 3A****LpTK3**

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAAACCCCTCTCGAAAGTT  
GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT  
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

**FIGURE 3B****LpTK4**

GTTCAACCGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATTGAGAGTGACGAC  
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCCGAGAGTGGCACAAAACC  
ACACAAATGAGTGCCGC

**FIGURE 3C****LpTK13**

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC  
AGTGATTTTCGGACTTTCCAAAGCACTGCGTGCTGATGAAACTACTACAAGGCCCAGACC  
CATGGAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC  
AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

**FIGURE 3D**



1 TTCGAGCTCG CCGGACNTTG ATTATTGACT AGTATTATTAT AGTAATCAAT TACGGGGTCA TTAGTTTCATA GCCCATATAT GGAGTTCGCG GTTACATAAC  
AAGCTCGAGC GGCGTGTAAC TAATAACTGA TCANTAATTA TCATTIAGTTA ATGCCCCAGT ATCAAGTAT CGGTATATA CCTCANGGCG CAATGTATTG

101 TTACGGTAA TGGCCGCTT TGGTGGGCTT GGTGACGCG CCAACGACC CCGCCCATTTG ACCTCAATTA TGACGTATGT TCCATATGA ACCCAATAG GGACTTTCCA  
AATGCCATTT ACCGGGCGGA CCGACTGCG CCGACTGCG GGTTCCTGG GCGGGTAACT TCGATTAAT ACTGCATACA AGGTATCAT TCGGTATATC CCTGNAAGGT

201 TTGACGTCAA TGGGTGGAGT ATTTACGTA ACTGACCAC TTGGCAGTAC ATCAAGTGA TCATATGCCA AGTACGCCC CTATTCACGT CAATGACGGT  
AACTGCAGT ACCCACCTCA TAAATGCCAT TTGACGGGTG AACGTCATG TAGTTCACAT AGTATACGT TCATGCGGG GATAACTGCA GTTACTGCCA

301 AATGGGCGG CCTGGCATTA TGCCCACTAC ATGACCTTAT GGCACCTTCC TACTTGGCAG TACATCTACG TATTAGTCAI CCGTATTTACC ATGCTGATGC  
TTACCGGGC GACCGTAT ACAGGTCATG TACTGATATA CCTGAAAGG ATGAACCTTC ATGTAGATTC ATATATCATA CGGATATATG TACCACATCG

401 GCTTTTGGCA GTACATCAAT GGGGTGGAT AGCGTTTGA CTCACGGGA TTTCACATC TCACCCCAT TGACGTCAAT GGGAGTTGT TTTGGCACCA  
CCAAACCGT CATGTAGTTA CCGGCACCTA TCGCCAACT GAGTGCCCT AAGGTTGAG AGGTGGGTA ACTGCATTA CCTCAACA NAACGTTGT

501 AATCAACGG GACTTTCCA ATGTGCTAA CAACTCGCC CCATTGACG CCATGACG TTAGCCGCG AATGGGCG TAGGCTGA CCGTGGAGG TCTATATAG CAGAGCTCGT  
TTAGTTGCC CTGAACGTT TTACAGCAT GTTGAGGCG GGTACTCG GTTACCCGCT TTAGCCGCG ATCCGCAT GCCACCTCC AGATATATC GTCGACCA

601 TTAGTGAAC GTCAGATCC CTGGAGACG CATCACGCT GTTTGACCT CCAATAGA CACCGGACC GATCCAGCT CCGGCGCCG GAACGTTGCA  
AATCACTTG CAGCTAGC GACCTCTGCG GTAGTGCGA CAAACTGA GGTATCTTCT GTGCCCCG CTAGGTGCA GCGCGCCGCC CTGCCACGT

FIGURE 4A

5/35

```

701 TTGGAACGGG GATTCCCGCT GCCAAGAGTG ACGPAGTAC CGCCTATAGA GTCTATAGGC CCALITGGCT TCGTTAGAAC GCGCTACAA TTANTACATA
AACCTTGGCG CTAAGGGGCA CGGTTCACAC TGCATTTCATG GCGGATATCT CAGATATCCG GGTGAACCGA AGCANCTTIG CCGCGATGTT ANTIATGTAT

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GAATAACATC CACTTTGCCCT TTCTCTCCAC AGGTGTCAC TCCCAGGTCC AACTGCACCT
TGGANTACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTATTTGTAG GTGAACCGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTCTATC GATTGAATTC CCGGGGATC CTCFAGAGAT CCCTCGACCT CGAGATCCAT TGTGCTGGCG CGATTCTTT ATCACTGATA AGTTGGTGA
GCCAAGTAG CTAACCTTAA GGGCCCTAG GAGATCTTA GAGAGCTGGA GCTCTAGGA ACACGACCCG CCTAAGAA TACTGACTAT TCACCCACCT

1001 CATATTATGT TTATCAGTGA TAAAGTGTCA AGCATGACAA AGTGCAGCC GAATACAGTG ATCCGTGCCG CCTAGACCT GTTGANCGAG GTCGGCGTAG
GTATAATACA ATAGTCACT ATTTCACAGT TCGTACTGTT TCAACGCTCG CTATGTGAC TAGGCACGGC GGCATCTGGA CAACTTGCTC CAGCCGATC

1101 ACGGTCTGAC GACAGGCMAA CTGGCGGAA CTTGGGGGT TCAGAGGCCG GCGCTTACT GGCATTTCAG GAACNAGCG GCGTGTCTG ACGCACTGGC
TGCCAGACTG CTGTGCGTTT GACCGCCTTG CCAACCCCA AGTCGTGGC CGCGAATGA CCTGAAGTC CTGTTCGCC CCGACCGAGC TGGGTGACCG

1201 CGAAGCCATG CTGGCGGAGA ATCATAGCAC TTGGTGCCG AGAGCCGACG ACGACTGGCG CTCATTTCTG ACTGGGATG CCGCAGCTT CAGCAGGCG
GCTTCGGTAC GACCGCCTCT TAGTATCTG AAGCCACGGC TCTCGGCTGC TCGTACCCG CAGTAAAGAC TGACCTTAC GGGCGTGA GTCCTCCG

1301 CTGCTGCCCT ACCGCCAGCA CAATGGATCT CGAGGGATCT TCCATACCTA CCAGTTCTGC GCTGTGAGGT CGCGGCCGCA CTACTCTTG ATGATTTACT
GACGAGCGGA TGGCGGTCGT GTTACCTAGA CTTCCCTAGA AGGTATGAT GGTCAAGACG CCGACGTCCA GCGCGGCGT GATGAGAAC TACATATGA

```

FIGURE 4B

6/35

1401 CATATTACCA AGGAATAACT GCGGGGCACA GGGTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT GTACCGCTGG  
 GTATATGGT TCCTATTGA CCGCCCGTGT CCGAGTCCAC GACTTCCTG TATACACTTT CACTGGATCT TCCGTCTCC ACTCGGAGA CAGTCCGACC

1501 CATAAAGGGCC GCTTGAGGGC TCTTTGGTCA AGCAGTAACG CCAGTCTCTG GGAAGGCACC TGTTACTCAG CAGACCATGA AAGGGCTCT CCCTTTCTT  
 GTATTCCTGG CCACTCCCG AGAATCCAGT TCGTCAATGC GGTACAGAC CTTCCGTGG ACAATGAGTC GTCTGTACT TTCCCGAGA GGAAGAGGA

1601 GCAGCAGTCA GGAACACTC TGCTCCACCA GCTTCCTGTG GGAGCTGGA TATTATCCAG GCTGCCCCG AGTCATCCG AGCCCTAAC CCTCCCTGTG  
 CCTCCTCAGT CCTTGTGAG ACGAGGTGT CCAAGACAC CCTCCGACT ATANTAGTGC CGACGGGCG TCAGTAGGCC TCCGATTTG GAGGGACAC

1701 GTGCTTCACT GGTACACTC CTGTCTCACT TCAATGCTCC TCTTGGCTC CTGTTCTCTG TTGMACTTT GTATGAGATA GCAGAGAA TAGCGAAGT  
 CACGAAGTCA CCAGTGTGAG GACAGGTGA AATGACGGG AGAACGGGAG CACCAAGGAG AACCTTCAA CATCATCTAT CGTCTCTT ATGCTTTCA

1801 CTTAAGTCT TTGATCTTC TTATAGTGC AGAGAGAA TCTGAGCTA TGCTGCTTC TCTCTCTG CTTCAGCTAC CTGAAGCCG TTTCTTGTCT  
 GAATTCAGA AACTAGAAAG AATATTACG TCTCTCTT ACAGTGCAT ACGACGGAG AGAGAGAGC GAAGTCGATG GACTTCGGG AAGAGACAGA  
 349 O R P G S E O R

1901 ATACCTGCTC TCTATCTCT CACACTCCTC CGAGGCCAGC ACCATCCAC TGTCTGCTG GTGTCCACA GAGCTTTCT AGTCGTTG GGTATGGG  
 TATGACGAG AGATAGACGA GTGTGAGGAG GTCCGGTCTG TGTAGGTG ACAGACAGAC CACAGGTGT CTCGGAACA TCCAGCAACC CCAGTACCC  
 391 Y R S E I Q E C E E S A L V H G S D T Q N D V S G K Y T T P T H P

2001 AATTCCTCAA ATGCTCTCAT CTGGAGGAA CCAAGGGTCT CAGCCCTCT GGCAGGCAC CCGGGAAGG ACACCAATT GTATACCTG CCGCCACGC  
 TTAAGGAGTT TACAGAAGTA GACTCTCTT GTGCCCGA GTCCGGAGA CUGTCTCTG GGCCTTTCC TGTGGGTCAA CATATGGAC CGCCGTCG  
 308 P E E F T K H R S S G R T E A G R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGCGCTG CAGGCTTGGC GGGCTGTCTT CAGCCTCAGC CTGGGCGATG TGTAGGGCCA TGGTGGAGAC CTGGGAGAG CTGCCCTCTT CTGAGCTCTG  
 ACACCGCGAC GTCCGACCG CCCGACGCA GTCCGACTCG GACCGCTAC ACATCCCGT ACCACCTG GACGCTCTT GACGGAGAA GACTCGAGAC  
 215 S H R Q L S P P S D E A D A Q A I H L A M T S V Q S F S G E E S S Q

2201 AGAGCTGGCG GGGGCGATGC AGACTCTCTC TTCTCTTGC AGGCCCTTGC CCTGGAGCAG GTCCGCGAGG ATCTCCACCA GCTCCGAGAA TGCAGGTCTC  
 TCTCGACCG CCCGCTACG TCTGGAGGAG AAGGAGNAC TCCGGGCGC GACCTCTC CAGGGCTCC TAGAGTGT CGAGGCTCTT ACGTCCAGAC  
 241 S S R P A M C V E E E E O L G R G O L L D G L I E V L E S P A P R

2301 GCCTTGGGT CTCCGGACCA GCACTTCAGC ATGATGGGC GTATGGCGG AGTGGCCAGC TCCGGGGCCC TCATCTCTT GCCCTCTCTC AGCCGCTGGC  
 CGGAACCCCA GAGCCTGTG CCACTAGTCG TACTACGCG CATACGCCC TCACCGTTCG AGGCCCGG AGTAGGAA CA GGCAGAGAG TCGGCGACCG  
 208 A K P D G S W C N L M I R R I A P T A L E P A R M R T G D R L R Q

2401 AGACTCTCTC ATGATCTGC AGCCGAGGGT ACGGGAGGC CCCGAGAGC AGATCTCTC AGAGAGCAGC CCCAAGGAG CACACGTCTC TCTGGCTGTG  
 TCTGAGGAG TACTAGAC TGGGTCCCA TGCCTCTCG Y P S A G L S P I E W L L V G P S W V D S Q T T

175 C F E E N I Q V G P Y P S A G L S P I E W L L V G P S W V D S Q T T

2501 GTACACCTTG TCGAGATGC TTTCAGGGC CATCCACTTC AGGGCAGCC GGGCACTGCC CTTCGGGAGC TAGTCGGGT CTTCGTAGAT GTCCCGGCA  
 CATGTGGAAC AGCTCTACG AAGTCCCG GTAGGTGAG TCCCGTCCG CCCGTGACG GAACGCTGC ATCAGCCCA Y D P D K Y I D R A

141 Y V K D F Y S E P A M W K L P L R A S G K R V Y D P D K Y I D R A

2601 AGGCAAGT CACAGATCTT CACCAGTCT CTTTCGACA GCAGATGTT CCGAGCAGCC AGGCTCTGT GGATGCACTT TCGGAGGCC AGGAATCCA  
 TCCGCTTCA GTCTTAGA GTGGTGCACC GAAGGCTGT CGCTTACA GGTCTCTCG TCCAGAGCA CCTAGGTGA AGCCCTTCCG TCCTTGAGGT  
 108 L G F D C I K V V D S E S L L I N R A A L D R H I C K R S A L P E

2701 TCCCTCTGCG CACCTGGAG CTGTAGCAGA CAAGATCTT CATGTCAGC GGGCTCAGCC ACAGTCTCT AGCTCTTGT TCTGGAGAA CCCGCTCTCG  
 AGGAGACCG GTGACCTTC GACATCTCT GTTCTAGAG GTACCACTCG CCCGACTCG TCTCAGGAG TCGAGAAC AGACTCTCTT GGGCGAGCG  
 75 M G R A V Q F S Y C V L D E M T L P S L W L D E A E Q D P S A R R A

FIGURE 4D

FIGURE 4E

2801' TCCGCCCTCG GTCTTCGAGA ACCGGCGGAA GAGGACCTTG TGGTGTCTCC CCGCGCGCT CCGATCCAGC CTGGCGAGCT CCACCATGGC GCGGAAGCGT  
 AGCGGGAGC CAGAAGCTCT TGGCGGCTT CTCTGGGAC AGCAGCGAGG GCGCGCGGA GGCTAGGTGG GACCGCTCGA GTGGTACCG CCGCTTCGCA  
 41' G G E T K S P R A F L V R D S S G P R R R D L R A L E V M A R F R  
 2901 CCGCGGTGCT CGGAGACTT CTCTGCGGA TCCAGGAAGC TGGCTCGAGG GCGCCAGTC GTCCGCGCA GAGGCGCTC CATTCGCCCG CCGCGCGCGG  
 GCGCGAGCA GCCCTGAA GAGGACGCTT ACCTGCTCG ACCGAGCTCC CCGCGGTGAG CAGCGCGCT CTCCGCGAG GTAGCGGGC GTAGCGGGC  
 B G R Q E P S K E  
 3001 CCGCGCGCAG CCGCGCGCT' CACCGCGCAG GGGCTCGGC CCGACTCTA GAGTCGACCT GCAGAGCTT GCGCGCATG GCGCAACTG TTTATTCAG  
 GCGGGCGTC CCGCGGCGA GTGGCTCTC CCGAGCGCG GCGTCTGAT CTCAGCTGA CCGCTTCGA CCGCGGTAC CCGGTTCGAC GCGGTTCGAC  
 3101 CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATT CACAAATAA GCATTTTTT CACTGCATC TAGTGTGCT TTGTCCAAAC TCATCAATCT  
 GAATATTACC AATGTTATT TCGTTATCT AGTGTATAA GTCTTATT CATAAAAAA GTACGTAAG ATCMACCA AACAGGTTG AGTAGTACA  
 3201 ATCTATCAT GTCTGGATCG ATCGGGAATT AATCGGCGC AGCACCATGG CCGCAATAA CCGTGTAAAG AGCACTTG TTAGGTACCT TCTGAGCGG  
 TAGAATAGTA CAGACCTAGC TAGCCCTTAA TTAGCGCGC TCGTGGTACC GGACTTTAT GCACTTTC TCCTTGAACC ATCCATGGA AGACTCGGC  
 3301' AAAGAACCA GCTGTGNAAG TGTGTCACTT AGGTGTGGA AATGCCCCAG GGTCCCCAG AGCAGAGCT ATGCAAGCA TGCATCTCA TTAGTCAGCA  
 TTTCTTGGTC GACACCTTAC ACACAGTCAA TCCACACCT TTCAGGGGTC CGAGGGTGG TCGTCTTCA TACGTTCTT ACGTAGAGT ATCAGTCTT  
 3401' ACCAGGTGT GAAAGTCCC AGGTCCCCA GCAGCGAAG GTATGCAAG CATGCTCTC AATTAGTCAG CAACCTACT CCGCGCCCTA ACTCGGCCCA  
 TGTCCACAC CTTICAGGG TCCGAGGGT CGTCTGCTT CATAGTTTC GTACGTAGAG TTATCAGTC GTTGGTACA GCGCGGGAT TCAGGCGGGT

9/35

FIGURE 4F

3501 TCUCGUCCTT AACTCGGCC AGTCCGCC ATTCCTCGCC CCATGGCTGA CTATTTT TTTATATGC AGAGGCCGAG GCGCCTCGG CCTCTAGCT  
 AGCGCGGGA TTGAGCGGG TCAAGCGGG TAAAGCGGG GGTACCGACT GATTAAAAA AATAAATACG TCTCCGGCTC GCGCGAGCC GGAGACTCGA

3601 ATTCCAGAG TAGTGAGAG GCTTTTGG AGGCTAGGC TTTCGAAA AGCTGTAAAC AGCTGGCAC TGGCGTGTG TTTACAACTT CGTGACTGGG  
 TAAGGTCTC ATCACTCTC CGAARAAACC TCCGATCCG AAACGTTTT TCGACAAATG TCGACCGTG ACCCGAGCA AATGTGCA GCACTGACCC

3701 AAAACCCCTG CGTACCCAA CTTAATCGC TTGCAGCACA TCCCCCTTC GCACTGTGC GTAATAGCGA AGAGGCCCG ACCGATCGCC CTTCCTCAACA  
 TTTTGGGACC GCAATGGGT GANTAGCG AACGTCTGT AGGGGGGAG CGGTGACCG CATATCGCT TCTCCGGCG TGGTAGCG GAAGGTTGT

3801 GTTGGGTAG CTGAATGGG AATGGGCTT GATGGGTAT TTCTCTTA CGATCTGTG CCGTATTCA CACCGATAC GTCAAGCAA CCATAGTACG  
 CAACCATCG GACTTACCG TTACCGCGA CTACGCCATA AAAGAGGAT GGTAGACAC GCATANAAT GTGGGTATG CAGTTCTGT GTATCATGC

3901 CGCCCTGTAG CGCGCATTA AGCGGCGCG GTGTGTGTG TACGCGAGC GTACCGCTA CACTTGCAG CGCCCTAGCG CCGCTCTT TCGTTCTT  
 GCGGCACATC GCGCGTAT TCGCGCGCC CACACCACA ATCGCGTGC CACTGGGAT GTGAACGGT GCGGATCG GCGGAGGA AGGNAAGAA

4001 CCCTTCTTT CTGCCCAGT TCGCGGCTT TCCCGTCAA GCTCPAATC GGGGCTCC TTATGGGTG CGATTTAGT GTTACGCA CCTCGACCC  
 CGGAAGGAA GAGCGGTGA AGCGCGGA AGGGCACT CGAGATTAG CCGCGAGG AATCCCAAG CTAATATCAC GAATGCCGT GGAGCTGGG

4101 AAAAATTC ATTTGGTGA TGGTCACT AGTGGGCAAT CGCCCTGATA GACGTTTTT CCGCTTTGA CGTGGAGTC CAGTTCTTT AATAGTGGAC  
 TTTTGTGAC TAAACCCACT ACCAATGCA TCACCGGTA GCGGACTAT CTTCAAAA GCGGAACT GCAACCTAG GTCAAGAA TTATCACCTG

10/35

FIGURE 4G

4201 TCTTGTCTCCA AACTGGGAACA ACACTCAACC CTATCTCCGG CTATCTCTTT GATTATTAAG GGATTTTGGC GATTTCGGCC TATGTGGTAA AANAAGAGCT  
AGAACAAAGGT TTGACCTTGT TGTGAGTTGG GATAGAGCCC GNTAGAAAA CTAATATTC CTAAGCCCG CTAAGCCCG ATAACCAAT TTTTACTCGA

4301 GATTTAACAA AATTTAAACG CGAATTTTAA CAAATATTA ACGTTTACAA TTTATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCGC ATAGTTAAGC  
CTAAATTTGTT TTTAAATTC GCTTAAATTC GTTTATATAT TGCMAATGTT AANAATACCAC GTGAGAGTCA TCTTAGACCA GACTACGGCC TATCAATTCG

4401 CAACCTCCCT ATCGCTACGT GACTGGGTCA TGGCTGGCC CGACACCCG CCAACACCCG CTGACCCGCC CTACACGGGT TGTCTGCTCC CGCATCCGC  
GTTGAGGCGA TNGCGATGCA CTGACCCCACT ACCGACCCG GCTGTGGC GCTGTGGC GACTGCCCGA ACAGACGAGG GCCGTAGCGC

4501 TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGTTTT CACCGTCATC ACCGAAACG CGGAGGCAGT ATTCTTGAG AGGAAGGC  
AATGCTGTG CGACACTGGC AGAGGCCCTC GACGTACACA GTCTCCAAA GTGGCAGTAG TGGCTTGGC CGCTCCGTCA TAAGNACTTC TGCTTTCCGC

4601 CTGCTGATAC GCTATTTTT ATAGGTTAA GTCATGATAA TAATGGTTTC TTACAGCTCA GGTGCCACTT TTGGGGAAA TGTGGCGGA ACCCTATTT  
GAGCACTATG CGGATAAAA TATCCANTTA CAGTACTATT ATTACCAAG ATCTGCAGT CCACCGTGA AGCCCCCTT ACACCGGCT TGGCGATAA

4701 GTTTATTTTT CTAAATACAT TCAATATGT ATCCGCTCAT ATCCGCTCAT GAGACATAA CCTGATATA TGCTTCATA ATATTGAAA AGGAAGATA TGAGTATCA  
CAAAATAAAA GATTATGTA AGTTATACA TAGGCGAGTA CTCTGTATZ GGGACTATTT ACCAAGTTAT TATAACTTTT TCCTTCTCAT ACTCATAGT

4801 ACATTTCCGT GTCCGCCCTTA TTCCCTTTTT TCGGGCATTT TCCCTTCTC TTTTGTCTCA CCCAGAACG CTGGGTAAG TAAAGATGC TGAAGATCAG  
TGTAAAGCA CAGCGGAAT AAGGAAAAA ACGCCGTAA ACGGAGGAC AAAAACCACT GGTCTTTGC GACCACTTC ATTTCTACG ACTTCTAGC

FIGURE 4H

4901 TTGGGTGCAC GAGTGGCTTA CATTGAACTG GATCTCAACA GCGGTAAGAT CATTGAGAGT TTTCGCCCCG AAGAACGTTT TCCANTGATG AGCATTFTT  
 AACCCACGCG CTCACCCCAAT GTAGCTTGAC CTAGAGTTGT CCCCATTCTA GGAACCTCA AAGCGGGCG TTCTGCANA AGGTACTAC TCGTGAANA

5001 AAGTCTCTCT ATGTGGCGG GTATTATCCC GTGATGACCG CGGGCAAGAG CAACCTCGTC GCCGATACA CTATCTCAG AATGACTGG TTGAGTACTC  
 TTCAAGACGA TACACCGCGC CATATATAGG CACTACTCG CCCCCTCTC GTTGAGCCAG CGGCTATGT GATAAGATC TACTGMACC AACTCATGAG

5101 ACCAGTCACA GAAAGGCATC TTACGATGG CATGACAGTA AGAGATTAT GCAGTCTGCG CATIACCATG AGTATACA CTGCGGCCA CTTACTTCTG  
 TGGTCAGTCT CTTTTCGTAG AATGCTTACC GTACTGTCT TCTCTTATA CGTCACGAG GTATTGGTAC TCACTATTGT GACGCGGTT GAATGAAGAC

5201 ACAACGATCG GAGGACCGNA CGAGCTAACC GCTTTTTCG ACACATGGG GGATCATGTA ACTGCGCTG ATCGTTGGA ACAGGACTG AATGAAGCCA  
 TGTTCGTAG CTCCTGGCTT CCTCGATTG CGAANAACG TGTGTATCC CTTAGTACAT TGACGGGAC TAGCAACCTT TGGCTCGAC TTTACTTCTG

5301 TACCAACGA CGAGCGTGAC ACCACGATGC CAGCAGCAAT GGCACAACG TTGCGCAAC TATTAACTGG CGAATCTATT ACTCTAGCTT CCGCGCAACA  
 ATGTTTGTCT GCTCGCACTG TGGTGCTAG GTCTGTCTTA CCGTGTGTC CCGTGTGTC AATATGACC GTTGTATGNA TGAGATGNA GGGCGTTGT

5401 ATTAATAGAC TGGATGGAGC CGATTAAGT TGCAGACCA CTCTGCGCT CGGCGCTTC GGCTGGCTGG TTATCTCTG ATAAATCTGG AGCGGTGAG  
 TAATATCTG ACCTACCTCC GCCTATTCA ACCTCTGCT GAAGACGGA CCGGGNAGG CCGACCGACC AATATAGAC TATTAGACC TCGCGCACTC

5501 CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGAG TCAGGCACT ATGATGAAC  
 GCACCCAGG CCCCATAGTA ACCTCTGAC CCGGTCTAC CATTCGGGAG GGCATAGCAT CATTAGTGT GCTGCCCTC ACTCCGTGA TACTACTTG



FIGURE 4I

5601 GAAATGACA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGCTAA CTGTGAGACC AGCTTACTC ATATATACTT TAGATTGATT TAAACTTT  
CTTTATCTGT CTAGCGACTC TATCCACGGA GTGACTAATT CGTAACCATI GACAGTCTGG TTCAANTGAG TATATATGAA ATCTAATCTAA ATTTTGAA

5701 TTTTAAATT AAAGGATCT AGGTGAAGAT CCTTTTIGAT AATCTCATGA CCAAAATCC TAACTGTGAG TTTTCTTCC ACTGAGGTC AGACCCCGTA  
AAAAATTAAT TTTTCTTAGA TCCACTTCTA GGAAAACTA TTAGAGTACT GCTTTAGG AATTGCATC AAAGCAAGG TGACTGCGAG TCTGGGCAI

5801 GAAAAATCA AAGATCTTC TTGAGATCTT TTTTCTGC GCGTAACTG CTGCTTGCAA ACAAATAAC CACCGCTACC ACCGCTGCTT TGTTCGCGG  
CTTTCTTAGI TTCTAGAG AACTCTAGGA AAAAAAGAG CGCATTAGAC GACGAACTT TGTTTTGTG GTGCGATGG TCGCCACCAA ACAAACGGCC

5901 ATCAAGAGCT ACCAATCTT TTTCGAGG TAACTGGCT CAGCAGAGCG CAGATACCA ATACTGTCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT  
TAGTCTCGA TCGTTGAG AAAGGCTTCC ATTGACCGAA GTCTGTCTCC GTCTATGCTT TATCAGGA AGTCACATC GCATCATC CCGTGTGAA

6001 CAAGAACTCT GTAGCACCGC CTACATACTT CGCTCTGCTA ATCTGTATC CAGTGGCTGC TGGCAGTGGC GATAAGTCT GTCTTACCGG GTTGGACTCA  
GTTCTTGAGA CATCGTGCGG GATCTATGGA GCGAGACGAT TAGGACATG GTCACCGAGG ACGGTACCG CTATTCAGCA CAGATGGCC CAACCTGAGT

6101 AGACGATAGT TACCGATTA GCGCGACGG TCGGGCTGAA CCGGGGCTC GTGCACACAG CCCAGCTTGG AGCGACGAC CTACACCGAA CTGAGATACC  
TCTGCTATCA ATGCGCTATT CCGGTCGCC AGCCCGACTT GCGCCCAAG CACGTGCTC GGTTCGAACC TCGCTGTCTG GATGTGGCTT GACTCTATGG

6201 TACAGCGTGA GCATTGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGCGG GACAGCTATC CCGTAAGCGG CAGGTCTGGA ACAGGAGAGC GCACGAGCGP  
ATGTGCACT CGTAATCTT TCGCGGTGCG AGGGCTTCC CTCTTCCGC CTGTCCATAG GCCATTCGCC GTCCAGCCT TGTCTCTCTG CGTGTCTCC

13/35

FIGURE 4J

6301 GCTTCCAGGG GGAACGCTT GGTATCTTAA TAGTCCTGTC GGGTTTCGCC ACGTGTGACT TGAGGCTCGA TTTTGTGAT GCTCGTCAGG GGGCGGAGC  
CGNAGGTCCC CCTTTGCCGA CCATAGAAAT ATCAGGACAG CCCAAGCGG TGCAGACTGA ACTGCGAGCT AAAACACTA CGAGCAGTCC CCCCCTCTG

6401 CTATCGAAA ACGCAGCAA CCGCGCTTT TTACGTTTC TGGCCTTTG CTGGCTTTT GTCACATCT TCTTCTCTG GTTATCCCTT GATTCCTG  
GATACCTTT TCGGTCGTT GCGCCGAAA ATGCCAAG ACCGGAAC GACCGAAG CCACTGTACA AGNAGGACG CAATAGGGA CTACACACC

6501 ATAAACCTAT TACCGCTTT GATGAGCTG ATACCGCTCG CCGCAGCGA ACGACGAGC AGTGACGAG GAGCGGAG ACGGCCAAT  
TATTGGCATA ATGCGGAAA CTCCTCGAC TATGCGAGC GCGTCTGCT TGTGCTCG CTTGCTCAG TCACCTGCTC CTTGCTCTC TCGCGGTTA

6601 ACGCAACCG CCTCTCCCG CCGTTGGTC GATTCATTAA TCCAGCTGC ACGACAGTT TCCGACTCG AAGCGGCA GTGAGCGCA CGCAATTAAT  
TGGTTTGGC GGAGAGGGC GCGCAACCG CTAGTAAT AGGTGACCG TGCTGTCCA AGGCTGACC TTTCGCTCT CACTCGCTT CGCTTAATTA

6701 GTGAGTACC TCACTCATTA GGCACCCAG GCTTACACT TTATGCTTC GGTCTGATG TTGTGTGAA TTGTGAGCG ATAACTAT CACACAGGA  
CACTCATGG AGTANGTAAT CCGTGGGTC CGAATGTGA AATACGAGG CCGAGCATAC AACACCTT AACCTGCTC TATTGTAAA GTGTCTCTT

6801 ACAGCTATGA CCATGATTAC GAATTA  
TGTGACT GGTACTAATG CTTAAT

FIGURE 5A

```

1  TTCGAGCTCG  CCGGACATTTG  AATTATTCACCT  AGTTATTTAAAT  TAGGGGGTCA  TTAGTTTCATA  GCCCATATAT  GGAGTTCCGC  GTTACATTAAC
   AACCTCGAGC  GGGCTGTAAAC  TAATAAACCTGA  TCAATTAATTA  ATGCCCCAGT  ATTCAAAGTAT  CCGGTATATA  CCTCMAAGCG  CATTGTATTG

101  TTACGGTAA  TGGCCCGGCTT  GGCTGACCGC  CCMAAGACCC  CCGCCCATTTG  ACGTCAATAA  TGACGTATGT  TCCCATAGTA  ACGCCAATAG  GGACTTTCCA
   AATGCCATTT  ACCGGGCGGA  CCGACTGGCG  GGTGTCTGGG  GCGGGTAAAC  TGCAGTTATT  ACTGCATACA  AGGTATTCAT  TCGCGTTATC  CCTGNAAGGT

201  TTGACGTCA  TGGGTGGAGT  AATTACGGTA  AACTGCCCCAC  TTGGCAGTAC  ATCAAGTGTA  TCATATGCCA  AGTAGGCCC  CTATTGACGT  CAAAGACGGT
   AACTGCAGTT  ACCCACCCTCA  TAAATGCCAT  TTGACGGGTG  AACCGTCAATG  TAGTTACAT  AGTATACGGT  TCATGCGGG  GATAACTGCA  GTTACTGCCA

301  AAATGGCCG  CCTGGCATTA  TGGCCAGTAC  ATGACCTTAT  GGGACTTTCC  TACTTGGCAG  TACATCTACG  TATTAGTCAT  CGCTATTACC  ATGGTGATGC
   TTTTACCGGG  GGACCGTAAT  ACGGTTCATG  TACTGGNATA  CCTGNAAGG  ATGMACCGTG  ATGTAGATGC  ATATCAGTA  GCGATAATCG  TACCACCTAG

401  GGTTTTGGC  GTACATCAAT  GGGCGTGGAT  AGCGTTTGA  CTCACGGGA  TTTCCAAGTC  TCCACCCCAT  TGACGTCAAT  GGGAGTTTGT  TTTGGCACCA
   CCMAACCGT  CATGTAGTTA  CCGGCACCTA  TCGCCMAACT  GAGTGCCCT  AAGGTTTCAG  AGGTGGGTA  ACTGCAGTTA  CCCTCNAACA  AAACCGTGCT

501  AAATCAACGG  GACTTTCCAA  AATGTGTA  CAACTCCGC  CCATTGACGC  AAATGGGCG  TAGGCGTGA  CCGTGGAGG  TCTATATAAG  CAGAGCTCGT
   TTTAGTTGCC  CTGAAGGTT  TTACAGCAAT  GTTGAGGCG  GGTAACTGCG  TTTACCCGCC  ATCCGCACAT  GCCACCCCTC  AGATATATTC  GTCTCGAGCA

601  TTAGTGAACC  GTCAGATCGC  CTGGAGACGC  CATCCACGCT  GTTTTGACCT  CCATAGMAGA  CACCGGGAC  GATCCAGCCT  CCGCGGCCCG  GAACGGTGCA
   AATCACTTGG  CAGTCTAGCG  GACCTCTGCG  GTAGTGCGA  CMAAAGTGA  GGTATCTTCT  GTGGCCCTCG  CTAGGTGCGA  GCGCGCCGCC  CTTGCCNCGT

701  TTGGMAACGG  GATTCCCGCT  GCCAAGAGTG  ACGTAAGTAC  CGCCTATAGA  GTCTATAGGC  CCACCTGGCT  TCGTTAGAAC  GCGGCTACAA  TTAATACATA
   AACCTTGGC  CTAGGGGCA  CGGTCTCAC  TGCATTTCATG  GCGGATATCT  CAGATATCCG  GGTGAACCGA  AGCAATCTTG  CGCCGATGTT  AATTATGTAT

```

15/35

FIGURE 5B

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GAATMACATC CACTTTGCCT TTCTCTCCAC AGGTGTCCAC TCCCAGGTCC AACTGCACCT  
TGGAAATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTTATTGTAG GTGNAACGGA AMGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCCGGGATC CTCTAGAGAT CCCTCGACCT CGAGTCGACT TTTTTTTTTT TTTTGTAGG CCAAAGGGTA CTCTTTTTTC  
GCCAAGATAG CTAACCTAAG GGGCCCCCTAG GAGATCTCTA GGGAGCTGGA GCTCAGCTGA AAAAAAANA AAAAACATCC GGTTCCTCAT GAAGNAANAAG

1001 TTTATTAAAT ACTCAGAGT CTAGGCCACA GCAATCTACT GTTCTCTCT CATTTTCTTA AACTATTTTG ATACCTATTT CTCAGACTTT ATGGGTATTT  
AAATAATTAA TGAGTCTTCA GATCCCGTGT CGTTAGATGA CAAGAGGAGA GTAAAGGAT TTGATAAAAC TATGGATAAA GAGTCTGANA TACCCGATNA

1101 AGACATTTCT CACATTTCCA TAGATAATAA CTCATCCGTT TTGCMCCCTG ATTCTCAATA TTAAGAGATT AAACTAATG TATATGACTC TCAGTTGACA  
TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGGCM AAGGTTGGAC TAAGAGTTAT AATTCTCTAA TTTTGATTAC ATATACTGAG AGTCMACTGT

1201 CATACCTGAG TACAGAAAA TTCCATCATT TCCTTCTGCA AATGAAAAA GACTTGGTTT TCTCAACAGC TGATCATTTT TTTTATGCAT AGAAAAAAT  
GTATGACTTC ATGTCTTTT AAGGTAGTAA AGGAGACGT TTTACTTTTT CTGAAGCANA AGAGTTGCG ACGTAGTAAA AAAATACGTA TCTTTTTTTA

1301 GTGCAATTAC TCCAGTACA ATCAAGTCAT TTAACATGGC TTTACCATCA TTGTAGTTAC AGGATATTTT AAAAGAGAAA AAAAAATCTC AAAGCACAGG  
CACGTTAATG AGGTTTCATGT TAGTTTCAGTA AATTGTACCG AANTGGTAGT AACATCAATG TCCATATAAA TTTTCTCTTT TTTTGTAGG TTTCTGTCTC

1401 TCCTGCTGTG CAGCAAGCA ATCAAAATCC TTCAATATAA CAGCCTGATG GGATTCAGCA ATCTGAGGAA TAATGAATA CCACCTTAAT CAGTAAACAG  
AGGACGACAC GTCGTTTCTGT TAGTTTNAAG AAGTATTATT GTCGGACTAC CCTAAGTGT TAGACTCCTT ATTACTTAT ATTACTTAT GGTGAGATTA GTCAATTGTC

1501 GAAATGCTA CAACAGTCAC TGAGTAAAA TTGCACTATC ATCTGTTGAT TCTCTGATC GACATTCAA ACAATAATG GAAATGTAAG TATCTCTTAA  
CTTTTACGAT GTTGTCACTG ACTCATTTT AACCTGATAG TAGAGAACTA AGAGAACTAG CTGTAAAGTT TGTATTATC CTTTACATTC ATAGACAAAT

16/35

FIGURE 5C

1601 AAGAGAAAT AACTTGGTTT AGTGTGCTTA ATTTTACCAG GCAGTGAGGA AATTATATAT CACCTTGAAT GTCTGCAGT GTTGCCCAAT CAATAAATG  
 TTTCTTTTAA TTGNAACCAA TCACACGAAT TAAATGGTC CGTCACTCCT TTAATATATA GTGGAAGTGA CAGGACGTCA CAACGGGTCA GTTATTTTAC

1701 CACAAATAAT CTTTTTCATA ATACATGGCC AACTTTATCC TATCACTTGA ATATGTCAGG ATAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG  
 GTGTTTATTA GAAAAAGTAT TATGTACCGG TTGAAATAGG ATAGTGAAT TATACAGTCC TATTTGACTA ACACGTCAAC CAACTATTGT AACATAAACC

1801 GAATGGAATTA TTGGAATTTG TTTTGCTACT TTATTATTTG ATATTCTTCT CCAGTGTTCA TCTTATGAAG TTATTGTCAT CTGAATATGA AGAGTCTGTT  
 CTTACCTAAT AACTTTAAC AAACGATGA AATAATTAAC TATAAGAAGA GGTCAACAAGT AGAATACTTC AATAAACGTA GACTTATACT TCTCAGACAA  
 506 O R I F N N A D S Y S S D T

1901 TCAAAATAGT CTTCAAGTTT CCAACGCAGT GTCTCAATG TAGGTGTTT CTTAGGCTCT GCATTCAGC ACTCCAACAT GATGTTGTAA AATTGCTGTG  
 AGTTTATCA GAAGTCAAA GGTGCGTCA CAGGTGTTTAC ATCCAGCAAG GAAATCCGAGA CGTAAGGTG TGAGGTGTGA CTACAACATT TTAACGACAC  
 492 E F Y D E L K W R L T E F T P R E K P E A N W C E L M I N Y F Q Q

2001 GACAGTTGGA TGGTTGCGGA AGTCTATAGT TTTGAGCCAA CATCTGGATT ACCTGGGCAC CTGTCAATACC ACTGTAAGC ATTTGCCCAT AAGTAATGAT  
 CTGTCAACCT ACCAAGCCT TCAGATATCA AACTCGGT GTAGACCTAA TGGACCCGTG GACAGTATGG TGACATTCOG TAAACGGTA TTCNTTACTA  
 459 P C N S P Q P L R Y N Q A L M Q I V Q A G T M G S Y P M K G Y T I I

2101 TTCATNAAGA AGGATTCCAA ATGACCATAC ATGGACTTA ATGCTGAATT TATTACTACG AATGGCTTCG GGGCAGTCC ACTTCACCGG CAGCTTTATT  
 AAGTATTTCT TCCTAAGGTT TACTGGTATG TAGCCTGAAT TACGACTTAA ATAAATGATC TTACCGAAGC CCGCGTCAGG TGAAATGGCC GTCGAATAA  
 425 E Y L L I G F S W V D S K I S F K N S R I A E P A T W K V P L K I

2201 TCGTGTCTAG ATTCTATAGT GTCTTCATTA TCTACCTTAA AACTCTGGC AAGTCCAAA TCTGCTACTT TGTAGATATT ATGTTACCA ACAGGAGCAT  
 AGCACAGATC TAAGTATCTA CAGAAATAT AGATGGAATT TTTGAGACCG TTCAGGTTTT AGACGATGAA ACATCTATAA TACAAGTGGT TGTCTCTGTA  
 392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGGCAGC CAGATCTCTG TGAATGTAGT TCCGAGACTC CAGATAGGCC ATTCAGAGG CAACCTGTGC CGCATGTCT ACCTGTTGAG TCAGATGGAT  
 AAGACCGTCG GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCGG TAAAGTCTCC GTTGGACACG GCGGTACAGA TGGACAACATC AGTCTACCTA  
 359 N R A A L D R H I Y N R S E L Y A M G S A V Q A A M D V Q Q T L H I

FIGURE 5D

2401 TTTTGAATCCA GTGTCAATTTT GGAGATATATTC TTGCAGACTTT CCATGTCTCA TCAACTCTGT AATAATATAA ATTGGATCTT CTAAGTGCA AACAGCATAA  
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y  
  
 2501 AGCTGGATAA GCTTTGGATG TCTTAGGTTT TTTTATATCT GTGCTCTCCCT CAGGAAGTCA TTTGGATCCA TTGAACCTGG TTTTAATGTT TTCACCTGCTA  
 292 L Q I L K P H R L N K M I Q A E R L F D N P D M S G P K L T K V A  
  
 2601 CTGGAGTGGT ATGTGTCAC AGACCTTCCC ATACTTCGCC AACTGACCA GATCCCAATC GCTTCAGAAG CTGTATGGAG TTGCGGTCTA TCTCCCATTTG  
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q  
  
 2701 GTCCACGGTT TTATACGACA AATCAATGG AGCTGGGACC TGGATCTTTA AGCATGGTTT CCCAGCTTG ACACACAGGC CGTCACTTGT CTTCGGTGTAG  
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y  
  
 2801 TGGCTCACAA ATTGTTTCAG TGTGAAAG ATTCTCTTC GGTGAGAA AATCCCTCT TCATCCAGTC TTTTAATTT GTAGTGTTT ACAACTGCTC  
 192 H S V F E N L T S F I R R R T L F F G G E D L R K I R Y H K V V A  
  
 2901 CATCTAAAC TGAAGAGAG AATCTCTCTT TTTGGCTTTC ACTTCTCTG ATTAGAAAG AACCGTCTT GTTTTCTGAA TATAATAGTT GTTCTCTGTC  
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A  
  
 3001 ATCTGATCTT CCGATTGCTC CAAAGAACCA CGGCTCTGCC TGTAGGCTTC TGTCTCAGC CACGTAGTTA GAAGGAATAT AGCCTTGTAG TTGCTGACTG  
 125 D S R G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S  
  
 3101 GAGCCATCTC GTCTTTTCTC CAAGTGCTG GCAMCCACC AGCCCTCATG CAAAGTGTC AGAATCTGAA GTTTGTACC TGCTCGGAAG CTCAAAGTCTC  
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

FIGURE 5E

3201 CAGCAGTCCG AGCCTGGTAA TCACAAAG CCACAAAGTA GTGCCATGC CTCCTGACT GGGGAGGCA AAGGCCCTT GGATTTTCAA TCACGGTTGA  
 GTCGTCAGGC TCGGACCAAT AGTTTGTTC GGTGTTTCAT CACCGGTACG GAGACACTGA CCCTCTCTGT TTCCCGGGA CCTAAAAGTT AGTCCCACT  
 59 E A T R A Q Y D F L A V F Y H G H R Q S Q P S C L A G P N E I V T S

3301 CTTGTCTGCC TCCGTGGACA AACAGGGAG ATAGGGTTCT AGGTACTCCC AGAGCTCTG ACAGATGTTG CTCATTGTGC CTTGCTGGG AGAAGAGGAG  
 GNACAGACGG AGGCACCTGT TTGTCCCTC TATCCCAAGA TCCATGAGGG TCTCGAGAC TGCTACAAAC GAGTAACAG GNACCACCCC TCTTCTCTC  
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

3401 CAGGGCTTCT CCTCTCCCC TTAGTCTCTG CGATCCACCT TATCTTCTTT CACCAGCAA CTTTGAAGTC AGCACCAACT CACCATACTT CGGAGAGTAT  
 GTCCCGNAGA GGGAGAGGG AATCAGAGAC GCTAGGTGGA ATAGAAGGNA GTGTCTCTT GAACTTCAG TCGTGGTTGA GTGGTATGAA GCCTCTCATA

3501 GCAAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTGCA GCAAGTCTTA CTTACCGGCT TGCTTCTGT GGCTGGAGGT GCTACCCCGA  
 CGTTTCAGGG CAAAGTCTAG TCAGGTCTG GACCCAAAGT CGTTCAGGAT GGACCTCTCT GAATGCCCCA AGGAAAGACA CCGACCTCCA CGATGGGGCT

3601 GGCRAAAGT AGCAGGAGCT GGGCAGCTGC TCACTAGGAA GGTGTCTTTT CTCTTATCT GCTTAAAGAT CCCACAACAA AAATAAATA AAATTAAAG  
 CCGTTTGTGAC TCGTCTCTGA CCCGTGAGC ACTGATCTT CCACAGAAA GNAGAATAGA CGAATCTTA GGGTGTGT TTTATTTTAT TTTAATTTTC

3701 GGCTTTATTT AGACAAATAT CTGAGNACAG AATGGTGCCA TCTTGCCCTT TGTCCTAATA AAAAGTTAGC AAGAGGAGC TACTAACCCC TGGTAAACCC  
 CCGAATATAA TCTGTTTATA GACTCTGTG TTAACCAAGT AGNACGGAAA ACAGGTTAT TTTCAATCG TTCTCTCTCG ATGATTGGGG ACCATTTTGG

3801 TCCACGTCTT GCTTTCGCCA GGTGCACTC GAGGATCTT CCATACCTAC CAGTCTGCG CGGCGCGGA CTCTAGAGTC GACCTGCAGA  
 AGGTGCAGAA CGAAGCGGT CCCAGCTGAG CTCCTTAGAA GGTATGGATG GTCAAGACGC GGACGTCCAG CGCCGGCGCT GAGATCTCAG CTGGACGTCT

3901 AGCTTGGCCG CCAATGGCCA ACTTGTAT TGCAGCTTAT AATGTTTACA AATAAAGCAA TACCATCACA AATTTCACA ATAAAGCAT TTTTTCACGT  
 TCGMACCGC GGTACCGGT TGMACNAATA ACGTCGAATA TTACCAATGT TTATTTCTGT TTAAGTGT TATTTCGTAA AAAAAGTGAC

19/35

FIGURE 5F

4001 CATTCCTAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG GATCGGGAAAT TAATTGCGCG CAGCACCATG GCCTGAAATA ACCTCTGAAA  
GTAAGATCAA CACCAACAG GTTTGAGTAG TTACATAGNA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCTGGGTAC CGGACTTTAT TGGAGACTTT

4101 GAGGAACCTG GTTAGGTACC TTCTGAGGCG GAAAGAACCA GCTGTGGAAAT GTGTGTCACT TAGGGTGTGG AAGTCCCA GGTCTCCCG CAGGCAGAAG  
CTCCTTGAC CAATCCATGG AAGACTCCGC CTTTCTTGGT CGACACCTTA CACACAGTCA ATCCACACC TTTCAGGGGT CCGAGGGGTC GTCCGCTTC

4201 TATGCAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCMAA GCATGCATCT CAATTAGTCA  
ATACGTTTCG TAGCTAGAGT TAATCAGTGG TTGGTCCACA CCTTTCAGGG GTCCGAGGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCCGCC CATTCCTCGC CCCATGGCTG ACTAATTTTT TTTATTTATG  
CGTTGGTATC AGGCGGGGA TTGAGGCGGG TAGGCGGGG ATTGAGGCGG GTCAAGGCGG GTAAGAGCGG GGTACCGAC TGATTAAAAA AATAAATATC

4401 CAGAGGCCGA GGCGCCCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCTTAGG CTTTTCGAAA AAGCTGTAA CAGCTTGGCA  
GTCTCCGGCT CCGCGGAGC CGGAGACTCG ATAAGGTCTT CATCACTCCT CCGAAAAAAC CTCCGGATCC GAAACGTTT TTCGACNAIT GTCGAACCGT

4501 CTGGCCGTCG TTTTACAAG TCGTGACTGG GAAACCCCTG GCGTTACCCA ACTTAATCGC CTTCAGCAGC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG  
GACCGGCAGC AATAATGTTG AGCACTGACC CTTTGGGAC CGCAATGGGT TGAATTAGCG GAACGTCTG TAGGGGAAA GCGGTCCGACC GCATTTATCGC

4601 AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGATGGC GATGGCGCC TGATGGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTT  
TTCTCCGGC GTGGCTAGCG GGAAGGGTTG TCAACGCGTC GGACTTACC GCTTACCGG CTTACCGCG ACTACGCCAT AAAAGAGGAA TCGGTAGACA CGCCATNANG

4701 ACACCGCATA CGTCAAAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGCG GGTGTGTGTGG TTACGCGCAG CGTGACCGCT AACTTGGCA  
TGTGGCGTAT GCAGTTTCGT TGGTATCATG CGCGGACAT CGCCGCGTAA TTCCGCGCCG CCACACCAACC AATCGCGTC GCACTGGCGA TGTGAACGGT



FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTGCGTTTCT TCCCTTCCTT TCTCGCCACG TTGCGCGGCT TTCCCGGCTCA AGCTCTAAAT CGGGGCTCC CTTTAGGGTT  
CGCGGGATCG CGGGCGAGGA AAGCGAAGA AGGCMGGA AGNGCGGTG AAGCGGCCGA AAGGGCAGT TCGAGATTTA GCGCCGAGG GAATCCCA

4901 CCGATTAGT GCTTTACGGC ACCTCGACCC CAAAAACTT GATTGGGTG ATGGTTACG TAGTGGCCA TCGCCCTGAT AGACGGTTT TCGCCCTTTG  
GGCTNAATCA CGAAATGCG TGGAGCTGG GTTTTGTGA CTAACCCAC TACCAAGTGC ATCACCCTT AGCGGACTA TCTGCCAAA AGCGGNAAC

5001 ACGTTGAGT CCAGTTCTT TAATAGTGA CTCTGTTC AACTGGAAC AACACTCAAC CCTATCTGG GCTATCTTT TGATTTATAA GGGATTTTGC  
TGCAACCTCA GGTGCAAGAA ATTATCACCT GAGAACNAGG TTTGACCTTG TTGTGAGTTG GGATAGAGCC CGATAAGAAA ACTAAATATT CCTNAAAACG

5101 CGATTTCGGC CTATTGGTTA AAAATGAGC TGATTTAACA AAAATTAAAC GCGAATTTTA ACAATATTT AACGTTTACA ATTTTATGTT GCATCTCAG  
GCTAAAGCCG GATAACCAAT TTTTACTCG ACTAAATTTG CGCTTAAAT TGTTTTATAA TTGCAATGT TAAATATCCA CGTGAGAGTC

5201 TACAATCTGC TCTGATCCG CATAGTTAAG CCAGCCCGA CACCCGCTGA CGCGCCCTGA CCGCTTGTG TGCTCCCGG ACAGGGGCG TAGGCGAATG  
ATGTTAGACG AGACTACGG GTATCAATTC GGTGCGGCT GTGGCGGTT GTGGCGACT GCGCGGACT GCGGAGGCG ACAGGGGCG TAGGCGAATG

5301 AGACAAGCTG TGACCGTCTC CGGAGCTGC ATGTGTCAGA GGTTTTCACC GTATCACC GACGAGCGA GACGAAAGG CCTCGTGATA CGCCTATTTT  
TCTGTTGAC ACTGGCAGG GCGCTCGACG TACACAGTCT CCAAAAGTGG CAGTAGTGG TTTGCGCGCT CTGCTTTCCC GGAGCACTAT GCGGATANA

5401 TATAGGTTAA TGTATGATA ATATGGTTT CTAGAGCTC AGGTGGACT TTTGCGGGA ATGTGCGCG AACCCCTATT TGTATTATTT TCTAAATACA  
ATATCCAAAT ACAGTACTAT TATTAGCAAA GAATCTGCAG TCCACCTGA AAAGCCCTT TACACGCGC TTGGGATANA ACNAATANA AGATTATGT

5501 TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATA ATGCTTCAAT AATATTGAAA AAGGAAGAT ATGAGTATTC AACATTTCCG TGTGCGCCTT  
ANGTTTATAC ATAGCGAGT ACTCTGTATT TGGNCTATT TACGAAGTTA TTATAACTTT TTCTTCTCA TACTCATAG TTGTAAAGGC ACAGCGGNA

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCTTTCCT GTTTTGTGTC ACCCAGAAAC GCTGGTGAAA GTAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT  
 TAAAGGAAAA AACGCCGTAA AACGCCGTAA AACGGAAGGA CAAAAGCGAG TGGGTCTTTG CGACCACCTTT CATTTCCTAGT CAACCCACGT GCTCACCCAA

5701 ACATCGAACT GGAATCTCAG AGCGGTNAGA TCCTTGAGAG TTTTTCGCCCC GAAGACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC  
 TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAACCTCTC AAAAGCGGG CTTCCTTGCAA AAGTTACTA CTCGTGAAA TTCAAGACG ATACACCCGG

5801 GGTATTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAGCAT  
 CCATAATAGG GCATAACTGC GCGCGTTCT CGTTGAGCCA GCGCGTATG TGATAAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG TCTTTTGGTA

5901 CTTACGGATG GCATGACAGT AAGAGAAATTA TGCAGTGTG TGCAATACCAT GAGTGATAAC ACTGCGCCA ACTTACTTCT GACAACGATC GGAGGACCGA  
 GAATGCTTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG TGACGCGGT TGAATGAAGA CTGTGTAG CCTCTGGCT

6001 AGGAGCTAAC CGCTTTTGTG CACAACATGG GGGATCATGT AACTCGCCTT GATCGTTGG AACCGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA  
 TCCTCGATTG GCGAAAAAC GTGTGTATAC CCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTGGA CTTACTTGG TATGGTTTC TCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAACT GTTGGCGAAA CTATTAACTG GGAACACTT TACTCTAGCT TCCCGGCAAC AATTAAATAGA CTGGATGGAG  
 GTGGTGTCTAC GGACATCGTT ACCGTGTGTTG CACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA AGGCGCTTG TTAATTATCT GACCTACCTC

6201 GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGCTGA GCGTGGTCT CCGGTATCA  
 CGCCTATTTC AACGTCTTGG TGAAGACGCG AGCCGGNAG GCCGACCGAC CAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAGCACT GGGGCCAGAT GGTAAAGCCCT CCCGTATCTT AGTTATCTAC ACCAGGGGA GTCAGGCAC TATGGATGAA CGAATAGAC AGATCGCTGA  
 AACGTCTGA CCCCCGTCTA CCATTCGGGA GGCATAGCA TCAATAGATG TGCTGCCCCCT CAGTCCGTTG ATACCTACTT GCTTTATCTG TCTAGCGACT

22/35

FIGURE 51

6401 GATAGGTGCC TCACTGATTGA AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT TTAGATTGAT TTAAACTTC ATTTTAAAT TAAAGGATC  
CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA GTATATATGA ATCTAACTA AATTTGAAG TAAAAATTAA ATTTTCCTAG

6501 TAGGTGAAGA TCCCTTTTGA TAATCTCATG ACCAAATCC CTTAACGTGA GTTTTCGTTT CACTGAGGT CAGACCCCGT AGAAAGATC AAAGGATCTT  
ATCCACTTCT AGGAAAACCT ATTAGAGTAC TGGTTTTAGG GAATGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGGCA TCTTTTCTAG TTCTCTAGAA

6601 CTTGAGATCC TTTTCTTCTG CCGTAATCT GCTGCTTCCA AACAAAAAA CCACCGCTAC CAGCGTGGT TTGTTTGGC GATCAAGAGC TACCAACTCT  
GAACCTCTAG AAAAAAAGAC GCGCATTTAGA CGACGAACGT TTGTTTCTT GGTGGCGATG GTGCCACCA AACAAACGGC CTAGTTCTCG ATGTTTGA

6701 TTTTCCGAAG GTAACCTGGT TCAGCAGAGC GCAGATACCA AATACTGTTT TTCTAGTCTA GCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG  
AAAAGGCTTC CATGACCGA AGTCGTCTCG CGTCTATGGT TTATGACAAG AAGATCACAT CCGCATCAAT CCGTGGTGA AGTTCTTGAG ACATCGTGGC

6801 CCTACATACC TGGCTCTGCT AATCTGTGTA CCAGTGGCTG CTGCCAGTGG CGATAAGTGG TGTCTTACCG GGTGGACTC AAGACGATAG TTACCGGATA  
GGATGTATAG AGCGAGACGA TTAGGACAAT GGTCAACCGAC GACGGTCACC GCTATTTCAGC ACAGATGGC CCAACCTGAG TTCTGCTATC AATGGCCTAT

6901 AGCGCCAGCG GTGCGGCTGA ACGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATAGAG  
TCCCGGTGCG CAGCCCGACT TGCCCCCAA GCACGTGTGT CCGGTGGAAC CTCGCTTGGT GGAJGTGGCT TGACTCTATG GATGTGGCAG TCGATACTCT

7001 AAGGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGCTCGG AACAGGAGAG CGCAGGAGGG AGCTTCCAGG GGGAAACGCC  
TTGCGGGTGC GAAGGGCTTC CTTCTTTTCCG CTTGTCCATA GGCATTTCCG CGTCCACGCC TTGTCTCTC GGTGCTCCC TCGAAGGTCC CCTTTTGGCG

7101 TGGTATCTTT ATAGTCTGCT CGGTTTTCGC CACCTCTGAC TTGAGCGTGG TTGCTCTCTC GCAGGAGAG CCGGCGGAG CCTATGAAA AACGCCAGCA  
ACCATAGAAA TATCAGGACA GCCCAAGCG GTGAGAGCTG AACTCGCAGC TAAAAACCT ACGAGCAGTC CCCCCGCTC GGATACCTTT TTGCGGTCTGT

23/35

FIGURE 5J

7201 ACGCGCCTT TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT  
TGCGCCGGAA AAATGCCNAG GACCGGAA ACGAGGTGAC AAGAAAGGAC GCAATAGGGG ACTATGGCAT CTATTGGCAT AATGGCGGAA

7301 TGAGTGAGCT GATACCGCTC GCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTCAGCGA GGAAGCGGAA GAGCGCCAA TACGCAAAACC GCCTCTCCCC  
ACTCACTCGA CTATGGCGAG CGCGTCGGC TTGCTGGCTC GGTCTGCTCA GTCACTCGCT CCTTCGCCCT CTGCGGGTT ATGCGTTTGG CGGAGAGGGG

7401 GCGCGTTGGC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCAT  
CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGCTGAC CTTTCGCCCG TCACTCGGT TCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCCTTACAC TTTATGCTTC CGCTCGTAT GTTGTTGGA ATTGTAGCG GATAACAATT TCACACAGGA AACAGCTATG ACATGATTAC  
TCCGTGGGT CCGAATGTG AATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA AGTGTGCTT TGTGCTATG TGTACTAATG

7601 GAATTAA  
CTTAATT

24/35

1 GCGGCCGCAG AGAAGCAGA GGATGGGGCT TAGCAGCTGG CAGAGCCAGG ACCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAAGNAGTC CTGAACCTTT  
CGCCGGGGTC TCTTTCGTCT CCTACCCCGA ATCGTCGACC GTCGCGGTCC TGGCCCCCTCC ATCGTCTTTC TGGTGTTCAT GTTCTTCAG GACTTTCANA

101 GGTTTTCTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAGACCC TGAAGATCAC CGACTTTGGC CTGSCCCGAG AGTGGCACAA AACCCACAA  
CCAAACGAC GACGTCGGGT AACTCTCACT GCTGTACCTC GTGTCTGGG ACTTCTAGTS GCTGAACCG GACCGGGCTC TCACCGTGT TTGGTGTGT

201 ATGAGTGCCG CXGGCACCTA CXCTGGATG GCTCCTGAGG TTATCAAGGC CTCCACCTTC TCTAAGGCCA GTGAGGTCTG GAGTTTGGG GTGCTGCTGT  
TACTCACGGC G?CCGTGGAT G?GGACCTAC CGAGGACTCC ATAGTTCCG GAGGTGGAAG AGATTCCCT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GCGAACTGCT GACCGGGGAG XTGCCATACC GTGGCATTGA CTGCCTTGCT GTGGCCTATG GCGTAGTCT TAACAGCTC ACACTGCCAT CCATCCACCT  
CCCTGACGA CTGGCCCCC ?ACGGTATGG CACCGTAAC GACGGACGA CACCGGATAC CGCATCGACA ATTGTTGAG TGTGACGGTA GGTAGGTGGA

401 GGCC  
CCGG

FIGURE 6

25/35

FIGURE 7A

1 ATGAGAGAGCT TGGGCGGCGA CCGCGCGAG CTGCGCTGCG TGGTGTGTTTT TTCTGCAAGG ATATTGCGA CTATTACAAA TCAAGATATCTG CTTGTGATCA  
TACTCTGCGA ACCGCGGCT GCGCGCGCTG GACGCGCGG GACGCGCGG AGCAATAAAA AAGAGCTTAC TATAAAGCTT GATAATGTTT ACTTCTAGAG CGACACTAGT

101 AGTGTGTTTT AATCAATGAT AAGAAAGATG ATTGATCAGT GGGAAATGA TTATCATATC CCATGCTATC AGATCCGCG GAGAGCGCTG GHTGTGCTTT  
TCACACAAAA TTAGTTAGTA TTCTTTTAC TAAGTACTGA GCGTTCACT ACTAATATAG GGTACCATAG TCTTAGGCGC CTCTGCGAGC CGACACGGAA

201 GAGACCCCG AGCTCAGGA CAGTGTACGA AGCTGCTGT CTGCAAGTGG ATGTATCTGC TTCCATCACA CTGCAAGTGC TGGTGGATGC CCCAGGGGAG  
CTCTGCGGTC TCGAGTCCCT GTACATGCT TCGACGCGGA CAGCTTCAGC TACATAGAGG AAGGTAGTGT GACGTTACG ACCAGCTAGC GGGTCCCTTG

301 ATTCTGTC TCTGCTCTT TAAACACAG TCGCTGAATT GGTAGCGAGA TTTGATTTA CAAGACAG GAGTTGTTTC CATGCTCAT TTCAAAATGA  
TAAAGACAG AGACCGAGAA ATTGCTGCG AGGAGCTTAA GGTGCTGCTT AAACATAAT GTTTGTGTC CTCACACAAAG GTACCGGTAA AACTTTTACT

401 CAGAAACCGA AGCTGAGAA TACCTACTTT TTATTACAG TCAAGCTAGC AATTAACAAA TATTGTTTAC ACTGACTATA AGAAATAGC TGGTTTACAG  
GTCTTTGGT TCGACCTCTT ATGATGAA AATAAGTCTC ACTTGATG TTAATGCTTT ATAACAATG TCACTCATAT TCTTTATGCG ACGAATGTC

501 ATTAAAGAG CTTTACTTTA CAAAATGGA AAGCCAGGAG GCGCTGGCT GCATATCTGA GAGCGTTCCA GAGCGGATCC TGGATGGGT GCTTTGGCAT  
TAATCTTCT GAAATGAAT CTTTTTACCT TTTGCTCTG GGGACGAGA CGTATAGACT CTGCGAGGT CTCGCTTAGG ACCTTACCCA CGAAACGCTA

601 TCACAGGGGG AAGCTGTAA AGAGTAAGT CGAGCTGTC TTAAAGGA GGAAGAGTC CTTGATGAAT TATTTGGAG CGACATAAG TGGTGTGCGA  
AGTGTCCGC TTGAGCATT TCTTCTTGA GGTGAGAAC AATTTTCT CTTTTTCA GAGTACTTA ATAAAGCTG CCTGTATTCC AGGACAGGT

701 GAAATGAAGT GGGAGGGA TCGACAGGC TGTTCACAAAT AGATTAAT CAACCTCTC AGACCACTT GGCACATTA TTTCTTAAAG TAGGGAGCC  
CTTAATTTA CCGTCCCTT AGTGTGCG AAGTCTTA TCTAGATTTA GTTGAAGAG TCTGTGTAA GGTGTTAAT AAGAAATTC ATCGCTTTC

FIGURE 7B

```

801 CTTATGAGATA AGGTGCAAAAG GTGTTTCATCT GAACCATGGA TTGCGGCTCA CCTGGAATT AGAACAACAA GCACCTCGAG AGGCAACTA CTTTGAGATG
GAATACCTAT TCCAGGTTTC GACAAGTACA CTTGGTACT AGCCCGAGT GACCCCTAA TCTTTGTGTT CGTGAGCTCC TCCCGTTGAT GAAACTCTAC

901 ACTACCTTATT CAACAACAG AACTATGATA CGGATCTCTT TTGCTTTTCT ATCATCTACT GCAAGAAACG ACACGGATA CTACACTTGT TCCTCTTCAA
TCATGATATA GTTGTGTC TTGATACTAT GCTTAAGACA AACAACAACA TACTAGTCAC CTTCTTTTC GATGTGAAC AGGAAGAAGTT

1001 AGCATCCGAG TCAATCAAGT TTGGTTAGCA TCGTAGAANA GGGATTATTA AATGCTACCA ATTCAAGTGA AGATTATGAA ATTGACCAAT ATGAAGAGTT
TCTTAGGCTC ACTTAGTCCA AACCAATGGT AGCATCTTTT CCTAATAT TTACGATGCT TAAGTTCACT TCTAATACTT TAAGTGTTA TACTTCTCAA

1101 TTCTTTTTCT GTCAGGTTTA AGCCCTACCC ACAATACAGA TGTACTGGA CTTCTCTCTG AATATCATTT CTTTGTUANC AAGGGTCT TCAATACGGA
AACAAAAACA CAGTCCAAAT TTGGAATGG TGTTAGTCT ACATGACCT GGAAGAGAGC TTTTAGTAAA GGAACACTCG TTTCCCAAC ACTATTGCT

1201 TACAGCATAT CCAAGTTTTG CAATCATAAG CACCACGAG CAGATATAT ATTCCATGCA GAATATGAG ATGCCCAATT TACCAAAATG TTCACGGTCT
ATTTCGTATA GTTCAAAAC GTTAGTATTC GTGTGCTGTC CTCTTATATA TAAGGTACTT CTTTACTAC TACGGTTAA ATGGTTTTAC AAGTGGACA

1301 ATATAGAAAG GAACCTCNA GTGCTCGCAG AAGCTTGCG AGCTAAGCG TCCGTCTCT CCGATGGATA CCAATTACCA TCTTGACCT GGAAGAAAGT
TATATTCTTC CTTTGAATT CAGGAGCTC TTGGAAGCG TTCACTGCG AGGACAAAGA CCTACCTAT GGTAAATGCT AGAACCTGGA CTTTCTTCC

1401 TTCAACAAAG TCTGCCAAT GCACAGAA GAACAGAA GAACTCTGA ATAGAAAGC TAACAGAAA GTGTTTTGAC AGTGGTCTC GAGCACTACT
AAGTCTGTC AGAGGTTCA CTTCTCTCT CTAGTCTCTT CTTAGTCTT TATCTTTTCG ATTGCTTTT CACAAACCTG TCACCCACAG CTCGTCTATG

1501 CTAAACATCA GTGAAGCCAT AAGAGGTTG CTGCTCAAGT GGTGTCATA CAATTCCTT GGCACATCTT GTGACAGGAT CTTTTAAAC TCTCCAGGCC
GATTTCTACT CACTTCGTA TTTTCCCAAG GACCACTTCA GCACAGGAT GTTAAGGAA CCGTGTAGA CACTCTGTA GGAATTTG AGAGGTCGG

```

FIGURE 7C

1601 CTTTCCTTT CATCCAGAC AACATCGAT TCTATGCAAG AATTGTTT TGTCTCTCT TCAATGTCTG TTTAACCTG CTAATTTGTC ACNAGTACNA  
 CGAAGGNAAG GTAGGTTCG TTTTAAGTA ACATAGTTC TTAACGAAA ACAGAGGNA AATTGGAC GATTAAACAG TGTTCATGTT

1701 AAAGCAATTT AGGTATGAAA GCGAGCTACA GATGGTACAG GTGACGGGAT CCTCAGATAA TGAGTACTTC TACGTTGATT TCAGAGATA TGAATATGAT  
 TTTCTTAAA TGCATACTTT GGTTCATCT CTAACATCT CACTGCTCTA TGGAGCTATT ACTCATGAG ATGCAACTAA AGTCTCTTAT ACTTATACTA

1801 CTCAAATGCG-AGTTTCCAG AGAAATTTTACAGTTTGGGA AGGTACTAGG ATCAGGTGCT TTTGGAAGG TGATGAAGG AACAGCTTAT GGAATTAGCA  
 GAGTTTACCC TCNAGGTTG TCTTTTAAAT CTCNAAACCT TCCATGATCC TAGTCCACGA AACCTTTTC ACTACTTGG TGTGGAATA CCTTAATGCT

1901 AACAGAGAGT-CTCAATCCAG GTTACCCTCA AATGTGTGA AAGAAAGCA GACAGCTCTG AACAGAGGG ACTCATGTCA GAACTCAAGA TCATGACCCG  
 TTTCTCTCA GAGTTAGGTC CAATGGCAGT TTTACGCACT TCTTTTCTGT CTGTGAGAC TTTCTCTCG TGAGTACAGT CTTGAGTTCT ACTACTGGGT

2001 GCTGGGAAGC CACGAGATA TTGTGAAGCT GGTGGGGGG TGACACTGT CAGAGCAAT TTACTTGATT TTGCAATCT GTTGTATCG TGATCTTCTG  
 CGACCTTCT GTCTCTTAT AACACTTGA CGACCCCGG ACCTGTGCA GTCTGTGTA ATGCACTAA AACCTATCA CAAGATACC ACTAGNAGG

2101 AACATCTAA-GAAGTAAAG AGAAATTT CACAGGACTT GCACAGAT TTTCAAGGA CACAATTCG GTTTTACCC CACTTCCAA TCACATCCAA  
 TTGATAGATT CTTCAATTTT TCTTTTAAA GTTCTCTGAA CTGTCTCTTA AAGTTCTCT AAGTAAAGT CAAAATGGG GTGAAAGGT AGTGTAGGT

2201 ATTCCAGCAT-GCCTGGTTCA AGAGAAGTTC-AGATACACC GCACTCGAT CAATCTCAG GCTTCATGG GAATTCATTT CACTCTGAG ATGAATTTGA  
 TAACTCGTA CGGACCAAGT TCTCTTCAG TCTATGTGG CCTGAGCTA GTTAGAGTC CCGAAGTACC CTTAAGTAA GTGAGACTTC TACTTTAAT

2301 ATATGAAAAC CAAAAAGGC TGGAGAAGA GAGGAGCTTG AATGTCTTA CATTGAGA TCTTCTTGG TTTGCATATC AGTTGCCAA AGGAATGGA  
 TATCTTTTG GTTTTTCGG ACCTCTCTCT CTTCTCTGAC TTACAGCAAT GTAACTCT GTAGATATAG TTCAACGGTT TCTTACCTT



FIGURE 7D

2401 TTTCTGGGAA TTAAGTCGTG TGTTCAACACA GACCTGGCCG CAGGAAAGCT GCTTGTCCACC CACGGGAAG TGGTGAAGAT ATGTGACCTT GGATTGGCTC  
 A AAGACCTTA ATTTCAGCAC ACAAGTGCT CTGGACCGC GGTCCCTTCA CCAACAGTGG GTGCCCTTC ACCACTTCTA TACACTGANA CCTAACCGAG

2501 GAGATATCAT GAGTGATTCC ACHATGTTG TCAGGGGGCA TUCCCGTGTG CCGTAAAT GGTGGCCCC CGAAGCCCTG TTGJAAGGCA TCTACACCAT  
 CTCTATGTG CTCACTAAGG TTGATACAA TTTTCCCTT ACGGGACAG CACATTTTA CCTACCGGG GCTTCCGAC AACTTCCGT AGATGTGGTA

2601 TAAGATGAT GTCTGGTCAT ATGCAATATT ACTGTGGCA ATCTTCTCAC TTGGTGTGAA TCCTTACCCCT GGCATTCCGG TTGATGCTAA CTTCTACAAA  
 ATTCTCACTA CAGACCAGTA TACCTTATTA TCACACCCCT TACAAAGACTG ACCACACTT AGGAATGGGA CCGTAAGGCC AACTAGCTT GAAGATGTTT

2701 CTGATTCAAA ATGATTTTAA ATGGATCAG CCATTTTATG CTACAGAGA AATATACATT ATAATGCCAA CTGCTGGGC TTTCGACTCA AGAAGCGGC  
 GACTAAGTTT TACCTAAATT TTACCTAGTC GGTAAATATC CATGCTCTTCT TTATATGTA TATTACGTTA GGACGACCCG AACTAGCTT TCCTTTGGCG

2801 CATECTTCCC TAATTGACT TCGTTTTAG GATGTCAGCT GGCAGATGCA GAAGAAGCGA TGTATCAGAA TGTGATGGC CGTCTTCCG AATGTCCTCA  
 GTAGCAAGG ATTAACTGA AGCAAAATC CTACAGTGA CCGTCTACGT CTTCCTCGCT ACATAGTCTT ACACCTACCG GCACAAAGCC TTACAGGACT

2901 CACCTACCAA AACAGCGGAC CTTTCAGCAG ACAGATGGAT TTGGGGCTAC TCTCTCCGCA GGCCTCAGCTC GAAGATTCTT AGAGAACAA TTTAGTTTTA  
 GTGGATGGTT TTGTCCGCTG GAAAGTCCTC TCTCTACCTA AACCCCGATG AGAGAGGCGT CCGAGTCCAG CTTCTNAGCA TCTCCTTGT AATCANAAT

3001 AGGACTTCAT CCGTCCACCT ATCCCTAACA GGCTGTAGAT TACCAACA AGTTAATTT CATCACTAA AGAATCTA TTATCACTG CTGCTTCACC  
 TCCCTGAAGTA GGGAGGTGGA TAGGATTTCT CCGACATCTA ATGGTTTTGT TCAATTAAG GTAGTCAATT TCTTTTAGAT AATAGTTGAC GACGAAGTGG

3101 AGACTTTCT CTAGAGAGCG  
 TCTGAANAAG CATCTCTCC

29/35

FIGURE 8A

1 TCGGCTCCA CCGCCCAAG GAGAGTCAGA CCTGKGGG GAGGGCCC CCAACTCAG TTGGATCCT ACCGAGTGA GCGCGGCCA TGGAGCTCC  
 AGCCGAGGT GCGCGGTCC CTCTCACTCT GACCCCCC GCTCCCGGG GATTGAGTC AAGCTAGGA TGGCTCACT CCGCGCGCT ACCTGAGGC  
 M E L R

101 GGTGCTGCTC TGTGGGCTT GGTGCGCC AGCTTTGGA GAGACCTGC TGAACACAA ATTGMAACT GCTGATCTGA AGTGGTGAC ATTCCCTCAG  
 CCAAGACGAG ACGACCGAA GCNACCGG TCAAACTTT TCTTGNGG ACTTGCTT TAACTTTGA CCACTAGACT TCACCCACTG TAAAGGAGTC  
 S V L L C W A S L A A A I E E T L L N T K L E T A D L K W V T F P Q

201 GTGACGGCC AGTGGGAGGA ACTGAGCGC CTGATGAGG MACAGCAG GTGCGCAC TACGAGTGT GTGACGTGA GGTGCCCC GCGCAGGCCC  
 CACTGCGG TCACCTCCT TGACTCGCG GACTACTCC TTGCTGTCTC GCACGCGTG ATGCTTCA CACTGCAGT GGCACCGGG CCGTCCCGG  
 J R V D G Q W E E L S G L D E E Q H S V R T Y E V C D V Q R A P G Q A H

301 ACTGGCTTGG CACAGGTGG GTCCACGRC GRRGRRGT CCAGGTGAC GCGAGCGTC GCTTACCAT GCTGAGTGC CTGTCCCTGC CTGCGGCTGG  
 TGACCGAAGC GTGTCCAACC CAGGTGCGT CCGCGGCA GTTCACTC CGGTCCGAG CGAGTGTA CGAGTCAAG GACAGGACG GAGCCGACC  
 72 W L R T G W V P R R G A V H V Y A T L R F T H L E C L S L P R A G

401 GCGTCTGTC AAGGAGACT TCACGTCTT CTACTATGAG AGGATGCG ACACGGCCAC GCGCTTCAAG CAGCCCTGA TGGAGAACCC CTACATCAAG  
 CCGAGACG TTCCTCTGA AGTGGCAGAA GATGATCTC TCGTACGCC TGTCCCGTG CCGGAGTGC GGTGCACT ACCTCTGGG GATGTAGTTC  
 105 R S C K E T F T V F Y Y E S D A D T A T A L T P A W H E N P Y I K

501 GTGGACAGG TGGCGCGGA GCATCTCAC CGGAGCGCC CTGGGCGGA GGGCACCGG AAGTGAAATG TCAGACGCT GGTCTGGGA CCGCTCAGCA  
 CACTGTGCC ACCGCGGCT GTAGAGTGG GCTTCGCG GACCCGGCT CCGTGGGCC TTCCACTTAC AGTTCTGGA CCGAGACCT GCGGAGTGT  
 139 V D T V A A E H L T R K R P G A E A T Q K V N V K T L R L Q P L S K

601 AGGTGGCTT CTACCTGCG TTCCAGGACC AGGTGCTG CATGGCCCTG CTATCCCTGC ACTCTTCTA CAAGAATGC GCCCAGCTGA CTGTGAACT  
 TCCGACCGA GATGACCGG AAGTCTCTG TCCACGGAC GTACCGGAC GATAGGACG TGGAGAGT GTTTTCAG CCGGTGAGT GACACTTGA  
 172 A G F Y L A F Q D Q G A C H A L L S L H L F Y K K C A Q L T V N L

30/35

FIGURE 8B

801 TACTGCTGAG AGGATGGCCA GTGGGCGGCA CAGCGCTCA GGGCTGCGAG CTGTGCTCCG GGGTTGAGG CAGGTAGGG GAAACACCAAG TGCCGAGCCT  
 ATGACGGCAC TCCTACCGGT CACCGCGCTT GTTGGTCAGT GCGCGACGTC GACACGAGGC CCCAGCTCC GTGCACTCC CTGTGTGTTT ACGGCTCGGA  
 218 Y C R E D G Q W A E Q P V T G C S C A P G F E A A E G N T K C R A C

901 GTGCCCCAGG CACCTTCAAG CCCCTGTCAG GAGAAAGCTC TGCCAGCCA ATAGCCACTC TAACACCAAT GATCAGCCG TCTGCCAGTG  
 CACGGGTCCC GTGGAAGTTC GGGGACATTC CTATTGTCAG GACGGTCCGT ACGGTTCGGT TATCGGTGAG ATTGTGTAA CCTAGTGGC AGACGCTCAC  
 272 A Q G T F K P L S G E G S C Q P C P A N S H S N T I G S A V C Q C

1001 CCGGCTGGG TACTTCCGG CACGCACAGA CCCCCTGGT GCACCTGCA CCACCCCTCC TTGCGCTCG GAGGCGTGG TTTCCTGCTT GAAAGGCTCC  
 GACGAGCCC ATGAAAGGCC GTGCGTGTCT GGGGCCCCA CCGTGGACGT GGTGGGAGG AAGCCGAGG GCCTGCGACC AAAGGCGGA CTTCGCGAGG  
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S R L N G S

1101 TCCCTGCACC TGGAAATGAG TGCCCCCTG GAGTCTGGT GCCGAGAGA CCTCACCCTAC GCGCTCCCT GCGGAGTGG CCGACCCGGA GCGTCTGTTG  
 AGGAGGTGG ACCTTACCTC ACGGGGAC CTCACACCAAC CGGCTCTCT GAGTGTGATG CCGGAGCGGA CCGGCTCTAC GCGTGGGCT CCGAGGACAC  
 338 S L H L E W S A P L E S G G R E D L T Y A L R C R E C R P G G S C A

1201 CCGCCTGGG GGGAGACCTG ACTTTGACC CCGGCCCCG GGACTGGTG GAGCCCTGG GAGGCTACGT AGGCTACGT CCTGACTTCA CCTATACCTT  
 GCGGAGCCC CCGTCTGGAC TGAAMCTGG GCGCGGGGC CCGGACCAAC CTCGGGACCC ACCACCAAGC TCCCATGCA GACTGAAGT GGATATGGA  
 372 P C G G D L T P D P G P R D L V E P W V V V R G L R P D F T Y T F

1301 TGAGTCACT GCATTGAACG GGGTATCTC CTTAGCCACG GGGCCGCTC CATTTGAGCC TGTCATGTC ACCACTGACC GAGAGTACC TCGTGCAGTG  
 ACTCCAGTGA GGTAACTTGC CCCATAGGAG GAATCGTGC CCGGGCAGG GTAACTCGG ACAGTTACAG TGGTCACTGG CTCTCATGG AGGAGTCA  
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

1401 TCTGACATCC GGTGACCG GTCTCACCC AGACCTTGA GCGTGGCTG GCGTGTCCC CCGGCACCCA GTGGGCTGT GCTGACTAC GAGTCAANT  
 AGACTGTAG CCCACTGCG CAGGATGGG TCGTCAACT CCGACCGAC CCGACAAAGG GCGGCTGGT CACCCGACA CGACTGATG CTCCAGTTTA  
 438 S D I R V T R S S P S S L S L A W A V P R A P S G A V L D Y E V K Y

FIGURE 8C

1501 ACCATCAGAA GGGCGCCGAG GGTCCGACCA GGTTCGGTT CTTGAGAGG TCAGAAACC GGCAGAGCT GCGGGGGCTG AAGCGGGAG CCAAGCTACTT  
 TGTACTCTT CCGCGGCTC CCAAGGTGCT CTAATGCCAA GCACTCTGC LKKTSENR AEL RGL KRGA SYL  
 472 H E K G A E G P S S V R F L K T S E N R A E L R G L K R G A S Y L

1601 GGTTCAGCTA CGGCGCGCT CTAGCGCCG CTAGCGCCG TTGCGCCAG ACATCAGAG CAGACCCAA CTGGATGAGA GCGAGGCTG GCGGAGCAG  
 CCACTCCAT GCGCGCGGA GACTCCGCC GATGCCGAGG AAGCGGTCC TTGTAGTGC GTCTGGGT GACCTACTT CCGTCCGAC CGCTTCGTC  
 505 V Q V R A R S E A G Y G P F G Q E H H S Q T Q L D E S E G W R E Q

1701 CTGCGCCCTA TTGCGGCAC GGCAGTCTG GGTTCGCTC TGTTCCTGTT GGTCAATGTT GTGCGAGTTC TCTGCTCAG GAAACAGAGC AATGGAGAG  
 GACCGGACT AAGCGCGTG CCGTCAGCAC CCACACGAG ACCAGGACCA CCAGTAACAC CAGCTCAG AGACGAGTC CTTCGCTCG TTACCCCTC  
 538 L A L I A G T A V V G V V L V L V V L V A V L C L R K Q S N G R E

1801 AAGCAGATA TTGGACAA CAGCAGCT ATCTATCGG ACATGCTACT AAGTCTACA TCGACCCCTT CACTTATGAA GACCTAATG AGGCTGTGAG  
 TTCTCTTAT AAGCTGTCTT GTGCTGTCA TAGAGTAGCC TGTACCATGA TTCCAGATGT AGCTGGGGA GTCAACTT CTGGAATTAC TCCACACTC  
 572 A E Y S D K H G Q Y L I G H G T K V Y I D P F T Y E D P N E A V R

1901 GGAATTTGA AAGAGATCG ATGTCTCTTA CGTACAGAT GAGAGGTGA TTGTGCGAG TCAAGTTTGGC GAGGTGTGCC GGGGGGGCT CAAGGCCCA  
 CCTTAAGGT TTCTCTAGC TACAGAGGAT GCAGTTCTAA CTCTCCACT AACCAAGTCC ACTCAACGG CTCACACGG CCGCGCGGA GTTCGGGGT  
 605 E F A K E I D V S Y V K I E E V I G A G E F G E V C R G R L K A P

2001 GCGAAGAGG AGAGCTGTG GGCATCAG ACCCTGAAG GTGGCTACG GGAGCGGAG CGGCTGAGT TTCTGAGCA GCGCTCCATC ATGGCCAGT  
 CCTTCTTCC TCTCGACACA CCGTTAGTTC TGGGACTTCC CACCGATGT CTTGCGCTC GCGCACTCA AAGACTCGT CCGGAGGTAG TACCGGTCA  
 638 G K K E S C V A I K T L K G G Y T E R Q R R E F L S E A S I M G Q F

2101 TCAGGACCC CAATATCAT CCGCTGAGG GGTGGTCA CAAAGCATG CCGGTATGA TTCTACAGA GTTCATGAG AAGCGGCCC TGGACTCTT  
 AGCTGTGG GTTATAGTAG GCGGACCTCC CCGACCATG GTTGTGTAC GGGGAGTACT AAGAGTGTCT CAAGTACTC TTGCGCGGG ACCTAGGAA  
 672 F H P N I I R L E G V V T N S M P V M I L T E F M E N G A L D S F

32/35

FIGURE 8D

2201 CCTCGGCTA AACGAGGAC AGTTACAGT CATCCAGTC GTGGCATGC TCGGGGCAT CGCTCGGC ATCGGTACC TTCCGAGAT GACCTACTC  
 705 L R L N D G Q F T V I Q L V G M L R G I A S G M R Y L A E M S Y V  
  
 2301 CACCGAGCC TGGTCTCTG CACATCTTA GTCAACGCA ACCTGCTCT CAAATGTCT GACTTTGCG TTTCGGAAT CCTGAGGAG AACTCTTCC  
 738 H R D L A A R N I L V N S N L V C K V S D F G L S R P L E E N S S D  
  
 2401 ATCCACCTA CACGAGCTC CTGGAGGAA AGATTCCCAT CCGATGGACT GCGCGGAGG CCATTGCTT CCGAAGTTC ACTTCGGCA GTGATGCTG  
 772 P T Y T S S L G G K I P I R W T A P E A I A F R K F T S A S D A W  
  
 2501 GAGTTACGG ATTGTGATG GGGAGGTGAT GTCTTTTGG GAGAGCCGT ACTGGACAT GAGCAATCAG GACGTGATCA ATGCCATTGA ACAGGACTAC  
 805 S Y G I V M W E V H S F G E R P Y W D H S N Q D V I N A I E Q D Y  
  
 2601 CGGCTGCCC CGCCCCAGA CTGTCCCACC TCCCTCCACC AGTCATGCT GACTGTGAG GCAATGACC GGAATGCCG GCGCGCTTC CCGCAGGTG  
 838 R L P P P P D C P T S L H Q L M L D C W Q K D R N A R P R P P Q V V  
  
 2701 TCAGCGCCT GGCACAGATG ATCCGAGCC CCGCAGCCT CAAATCTG TCGCGGAGA ATGGCGGCG CTCACACCT CTCTGGACC AGCGGAGCC  
 872 S A L D K M I R N P A S L K I V A R E N G G A S H P L L D Q R Q P  
  
 2801 TCAGTACTCA GCTTTTGCT CTGTGGCGA GTGGCTTCG GGCATCAAAA TGGGAGATA CGAAGAAAT TTCCAGGCG CTGGCTTTGG CTCTTTCGAG  
 905 H Y S A F G S V G E W L R A I K H G R Y E E S F A A A G F G S F E

FIGURE 8E

2901 CTGCTCAGCC AGATCTCTGC TGAGGACTTG CTCGGAATCG GATTCACCTCT GCGGACAC CAGAGAAA TCTTGCCAG TGTCCAGCAC ATGAGTCCC  
 GACCACTCG TCTAGAGACG ACTCTGAC GAGCTTAG CTCAGTAGA CCGCCCTG GTCTCTTT AGAACGGTC ACAGTCTG TACTTCAGGG  
 938 L V S Q I S A E D L L R I G V T L A G H Q K I L A S V Q H M K S Q

3001 AGCCCAAGCC GGNACCCCG GTGGGACAG GAGACCCGC CCGCAGTAC TGACCTGCG GAACTCCCA CCGCAGGAC ACCCCCTCC CATTTTCGG  
 TCCGTTCCG CCGTTGGGC CCACCTGTC CTCCTGGCG GCGCTCATG ACTGACGTC CTGAGGGT GCGTCCCTG TCGCGAGG GTAAAGGCC  
 972 A K P G T P G G T G G P A P Q Y O P A G T P H P R D T A S P F S G

3101 GGCAGAGTGG GGACTCACAG AGGCCCCAG CCGTGTGCC CCGTGATG CACTTGAGC CCGTGGGTG AGGAGTTGC AATTTGAGA GACAGATTT  
 CCGTCTACC CCGAGTGTG TCCGGGGTC GGGACACGG GGCACCTAC GTGAACCTG GGCACCCAC TCTCAACG TTAACCTCT CTGTCTAAA  
 1005 A E W G L T E A P S P V P R W I A L O A R G V R S W Q F G E T G F

3201 GCGGTTCTG CCATMATAG AGGGANAT CACCCCCAG CCACCTGGR GAATCCAGA CCAGGGTA GCGGCCCTT CCGTCAGGAC TGGGTGTGAC  
 CCCCCAGAC GGTATATCC TCCCTTTTA GTGGGGTC GTGGAGCCC CTGAGGTCT GTTCCCACT CCGCGGANA GCGAGTCTG ACCACACTG  
 1038 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

3301 CAGAGGAAA GGAAGTCCC AACATCTCC AGCTCCCA GGTGCCCC TCACCTGAT GGTGGGTTC CCGCAGCA AGAGAGTGT GACTCCCTTG  
 GTCTCCTTT CTTACAGGG TTGTAGAGG TCGAGGGT CCACGGGG AGTCGACTA CCGACGAG GCGTCTGCT TCTCTACA CTGAGGAGC  
 1072 E E K E V P N I S Q P P Q V P P S P O W V R S R R P K R V O L P C

3401 CCAGTCCAG AGTGGGGGG CTGTCCACG GGCNAGAG GGTGTGAG GCCAGTGAC AAATCATTT GGGTTGTAG TCCCACTTG CTGCTGTAC  
 GGTGAGTTC TCACCCCCC GACAGGGTCC CCGTTCTTC CCCACAGTCC CCGGTACTG TTTAGTAC CCGAACATC AGGTTGAC GACGACGTG  
 1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C C H

3501 CACCAACTC ATCATTTTT TTCCCTTCTA ATGCCCTC CCGCAGTGC TCGCTCATA TTGAGTTT TCGAGTTTG TTTTGTCT TAATTTTCT  
 GTGTTGTAG TTAGTAAAAA AAGGGACAT TTACGGAG GGGTGGAG ACGGAGTAT AACTTCCAA AACTCAAAAC AAAAACCAGA ATTAAAAAGA  
 1138 H Q T Q S F F S L V N A P P P A A A F I L K V F E F C F W S O F F S

34/35

FIGURE 8F

3601 CCCCGTTCCC TTTTGTGTTTC TTGCTTTTGT TTTCTACCG TCCTTGTGAT AACTTGTGT TGGAGGGAAC CTGTTTCACT ATGGCTCTCT TTGCCCCAAGT  
GGGGCAAGG AANAACAAAG AAGCNAACA AANAGATGG AGGACAGTA TTGAACACA ACCTCCCTTG GACAAAGTA TACCGAGGA ACGGTTTCA  
1172 P F P F C F F V L F P Y R P C H N P V L E G T C P T H A S F A Q V

3701 TGAACAGGG GCCCATATC ATGTCGTTT CCAGAACAGT GCCTTGCTCA TCCACATCC CGGACCCCG CCTGGACCC CCNAGCTGT TCCTATGAAG  
ACTTTGTCCT CCGGTAGTAG TACAGACAAA GGTCTTGCTA CGGACCAAGT AGGCTGTAGG GGCCTGGGG GCACCTGGG GGTTCGACAC AGGATACTTC  
1205 E T G A H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

3801 CGGTGTGGG TGAGGTAGT AAGGGGGG TAGTGTGG TGGAACCCAG AACGGACCC CGGTGCTTGG AGGGTTCTT AATTATATT TAAAAAGTA  
CCACACCCC ACTCCATCAC TTTTCCCGC ATCAACCAAC ACCTTGCTC TTGCTGCG GCCAGAAC TCCCAAGNA TTTAATATA ATTTTTCAT  
1238 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

3901 ACTTTTGT TAAATAAAG AAATGGAC GTGTCCAGC TCCAGGGGTA AAAAAAAA AAAAAAAA  
TGAANAACAT ATTATTTC TTTTACCTG CACAGGTG AGTCCCAT TTTTTTTT TTTTTTT  
1272 F L Y K O K K H G R V P A P G V K K K K K

35/35

FIGURE 9                   ARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPMRWTAPEAIQYRKFFASAS

FIGURE 10                NVLVKSPNHVKITDFGLARLLEGEKEYNADGGKMPIKWMALCECIHYRKFTHQ

FIGURE 11                NCMLAGDMTVCVADFGLSWKIYSGATIVRGCASKLPVKWLALGSLADNLYTVHS

FIGURE 12                NCLVGKNYTIKIADFGMSRNLVSGDYY

FIGURE 13                TRNILVENENRVKIGDFGLTKVLPQDKEYYKVKEPGESPIFWYAPESLTSLSFVASD

FIGURE 14                ARNILVNSNLVCKVSDFGMSRVLEDDPEAAVYTRGGKIPIRWTAPEAIAYRKFTSASD



## INTERNATIONAL SEARCH REPORT

PCT/US 93/00586

International Application No

|  |  |   |
|--|--|---|
| <b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) <sup>6</sup>  |  |   |
| According to International Patent Classification (IPC) or to both National Classification and IPC  |  |   |
| Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68,<br>C12N15/11   |  |   |
| <b>II. FIELDS SEARCHED</b>   |  |   |
| Minimum Documentation Searched <sup>7</sup>  |  |   |
| Classification System  | Classification Symbols   |   |
| Int.Cl. 5  | C12N ; C12Q ; C07K   |   |
| Documentation Searched other than Minimum Documentation<br>to the Extent that such Documents are Included in the Fields Searched <sup>8</sup>  |  |   |
| <b>III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup></b>  |  |   |
| Category <sup>10</sup>   | Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>   | Relevant to Claim No. <sup>13</sup>                 |
| A  | <p>NEURON<br/>vol. 6, no. 5, May 1991,<br/>pages 691 - 704<br/>LAI, C. &amp; LEMKE, G. 'An extended family of<br/>protein-tyrosine kinase genes<br/>differentially expressed in the<br/>vertebrate nervous system'<br/>see the whole document</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p> | 1-7   |
| <p><sup>10</sup> Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p> |  |   |
| <b>IV. CERTIFICATION</b>   |  |   |
| Date of the Actual Completion of the International Search  |  | Date of Mailing of this International Search Report |
| 07 JUNE 1993   |  | 02 -07- 1993  |
| International Searching Authority  |  | Signature of Authorized Officer                     |
| EUROPEAN PATENT OFFICE   |  | ANDRES S.M.   |

| III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET) |  |                        |
|--|--|------------------------|
| Category *   | Citation of Document, with indication, where appropriate, of the relevant passages   | Relevant to Claims No. |
| A  | <p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA<br/>vol. 88, no. 23, 1 December 1991, WASHINGTON US<br/>pages 10411 - 10415<br/>HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p> <p>---</p> | 1-7,<br>13-15          |
| A  | <p>GENE<br/>vol. 110, no. 2, 15 January 1992, AMSTERDAM NL<br/>pages 205 - 211<br/>BRÄUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases'<br/>see the whole document</p> <p>---</p>  | 1-7,<br>13-15          |
| A  | <p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA<br/>vol. 87, November 1990, WASHINGTON US<br/>pages 8913 - 8917<br/>PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells'<br/>see especially clone JTK10</p> <p>---</p>  | 1-7                    |
| A  | <p>MOLECULAR AND CELLULAR BIOLOGY<br/>vol. 9, no. 4, April 1989, WASHINGTON US<br/>pages 1587 - 1593<br/>HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene'<br/>see the whole document</p> <p>---</p>   | 1-7,<br>16-18          |
| P,A  | <p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY)<br/>3 September 1992<br/>see the whole document</p> <p>-----</p>   | 1-12                   |

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 93/00586

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
obscurities.  
see additional page
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

OBSCURITIES

Claims 2,3,5 and 7 as filed in the application are in contradiction with the description. Indeed, they are dependent on claims 1,4 or 6 which are related to DNAs or proteins of human MEGAKARYOCYTIC origin, whereas they claim some sequences which are not from such an origin, if taking reference to page 8 of the description. Therefore the search of these claims has been made independently of claims 1,4 and 6.

Claims 14 and 15 relating to the nucleotide and the protein sequences of SAL-D4 are filed as being dependent on claim 10 which relates to SAL-S1. They have been interpreted and searched as depending on claim 13.

Claims 17 and 18 relating to the nucleotide and the protein sequences of LpTK 3 are filed as being dependent on claim 14 which relates to SAL-D4. They have been interpreted and searched as depending on claim 16.

Claims 20 and 21 relating to a DNA expression vector and a cell transformed therewith, are filed as being dependent on claim 17 which relates to a protein (LpTK 3). They have been searched as depending on claim 19.

Finally, claim 2, point a) relating to SAL-S1 refers erroneously to SEQ ID 7 which describes SAL-D4. It has been interpreted as referring on SEQ ID 5 and 17).

US 9300586  
SA 69794

**07/06/93**

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|---|---------------------|----------------------------|---------------------|
| WO-A-9214748                              | 03-09-92            | EP-A- 0536350              | 14-04-93            |
| -----                                     |                     |                            |                     |